

nature



NANOTUBES UNZIPPED

A route to graphene
nanoribbon electronics

RIISING SEA LEVELS
A fossil record

BIG BANG COSMOLOGY
Big bang of inflation

EVOLUTIONARY THEORY
Darwin on the wing

NATUREJOBS
Go with the wind

Abstractions



FIRST AUTHOR

Some consequences of climate change are already unfolding. Glaciers and ice sheets are melting, and sea levels are rising as a result. However, scientists aren't certain by how much the

rate of sea-level rise might accelerate; current predictions for increases until 2100 range from 0.3 centimetres to 1.4 centimetres per year. But Paul Blanchon, a geoscientist at the National Autonomous University of Mexico in Cancún, and his colleagues have learned that a sudden, catastrophic increase of more than 5 centimetres per year over a 50-year stretch is possible. On page 881, they describe their discovery that a sea-level jump of 2–3 metres already happened about 121,000 years ago. Blanchon tells *Nature* how and why it could recur.

How did you find out that sea levels had risen so quickly in the past?

We were studying fossil reefs along the Yucatán peninsula in eastern Mexico, looking for interruptions in the reefs' development, when we found two reef crests. One crest was about three metres above the current sea level, the other six. Some event had clearly disrupted their growth, killing the lower reef first and, within 50 years, allowing the higher one to develop into territory that is now farther inland. One possible cause of such disruption is an earthquake, but we know the peninsula was stable in the reefs' lifetimes. The only other possibility is a rapid sea-level jump of two to three metres, which would essentially have drowned the lower reef.

Did you have to dive to the ocean floor to study the fossil reefs?

No, a theme park has been excavated in the middle of these reefs, which are on land south of Playa del Carmen. There's no other place in the world where reefs of this age are so exposed. From the excavations, we were able to reconstruct the reef's internal structure in three dimensions.

Were there any challenges involved with working in a theme park?

One key site was in the middle of a jaguar and puma exhibit. We had to get up at five in the morning, lower our ladders into the pit, do our studies and get out of there before the jaguars and pumas were let out.

What do your results mean for sea-level rises in the future?

This earlier ice-sheet collapse happened during an interglacial, when it was warm and there wasn't a lot of ice around — just as it is on Earth today. We're assuming rapid ice loss from an ice sheet produced the jump in sea level, because it's the only known process that could generate such a rapid increase. This could happen again. ■

MAKING THE PAPER

Susan Lea

Structure of meningitis protein with human 'coat' yields vaccine clues.

Although relatively uncommon, meningitis — inflammation of the membranes covering the brain and spinal cord — can kill within hours. Immediately after infection, one major bacterial culprit — *Neisseria meningitidis* — coats itself with a human protein so that immune cells no longer recognize it as an intruder. Using X-ray crystallography, Susan Lea of the University of Oxford, UK, and her colleagues now describe the interaction between the microbe and the human protein.

Christoph Tang at Imperial College London had previously discovered that *N. meningitidis* uses a protein on its surface, dubbed factor-H-binding protein, to grasp hold of the human protein factor H. "Chris contacted us hoping that we could help to characterize these interactions further," recalls Lea.

Factor H is part of the complement system, an arm of the immune system that attacks foreign bodies in the bloodstream. To prevent the system from targeting the body's own cells, factor H circulates in the bloodstream and binds to sugar molecules on human cells, flagging them as 'self'. *N. meningitidis* is one of several bacteria that have hijacked this mechanism by producing proteins that can also bind factor H.

Human factor H is a long molecule, made up of 20 domains strung together like beads on a string. Tang and Lea's groups determined that only two such 'beads' — numbers six and seven — are key to binding the bacterial protein. "This was good news because it meant we could look at the structure of just two domains," says Lea.

But although the researchers had no trouble purifying and crystallizing the two factor-H domains in complex with the bacterial binding protein, solving the complex's structure proved tricky. "Some crystals are well behaved and some aren't. This one wasn't," laughs Lea. Because the



X-ray diffraction data were not of sufficient quality to deploy one of the two classical phasing methods, which use X-ray and computational information to resolve molecular structures, Lea and her colleagues had to find new ways to combine the information from both methods.

On page 890, the hard-earned structure reveals that the bacterial protein folds in the middle. Each half comprises a sheet of amino acids twisted into a barrel shape, with the human factor H portion stacked on top. "The structure essentially looks like two mugs side by side with a croissant on top," says Lea.

Lea and her co-workers singled out two amino acids that seemed to be crucial to the two proteins' tight association. When they mutated the two amino acids, the bacterial protein no longer bound factor H. "Without the structural information we would have had to mutate every single amino acid in the protein to identify these two," says Lea.

Meningitis is usually caused by viral or bacterial infection. For bacterial meningitis, the more severe form of the disease, there are vaccines for all but one type: meningitis B. Two candidate meningitis-B vaccines that are currently in clinical trials both use factor-H-binding protein as part of the vaccine formulation. However, this latest work suggests that the bacterial protein will immediately become bound by factor H and so will not generate an optimal immune response. "The mutated bacterial protein we have produced could make a better vaccine candidate," Lea says, "because it doesn't bind factor H." ■

FROM THE BLOGOSPHERE

Glimpsing how a chemist views the world — both inside and outside of science — enlightens and inspires on The Sceptical Chymist, NPG's chemical-community blog. In the 'Reactions' series, *Nature Chemistry* associate editor Neil Withers probes the personalities behind the pipettes.

In the latest instalment, Jonathan Clayden, an organic

chemist at the University of Manchester, UK, shares why he switched from molecular biology during his first year at university (<http://tinyurl.com/cpp4rc>). "The way that mechanistic explanations apply equally in flasks and in cells intrigued me. I found the ... way you can write a structure on paper and then plan how to make the molecule in the lab very appealing too. I spend a lot

of my time now trying to bend those same mechanistic rules to see when they break."

And if he weren't a chemist? Clayden says he'd like to be a gentleman farmer, circa 1910, with a well-stocked vegetable garden for cooking. It seems he cannot completely escape his teenage fascination "with the way that complicated things grow when simple rules collide." ■

Visit *Nautilus* for regular news relevant to *Nature* authors ▶ <http://blogs.nature.com/nautilus> and see Peer-to-Peer for news for peer reviewers and about peer review ▶ <http://blogs.nature.com/peer-to-peer>.

John Maddox 1925–2009

In memory of a transformative editor of *Nature*.

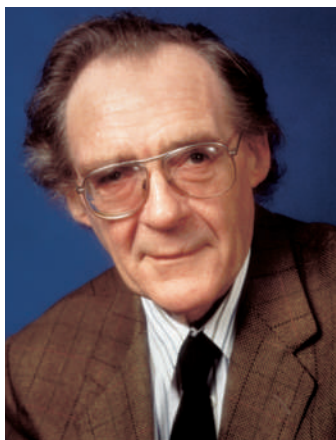
It was with great sadness that I and my colleagues at *Nature* learned of the death on Sunday of Sir John Maddox — or ‘JM’, as his colleagues always referred to him.

There was puzzlement, too. Yes, John had been looking frail recently, but, well, this was JM — the perpetually restless, irresistible, unstoppable force. The editor who conducted some gatherings with ‘shock and awe’ as some recall. The ‘man with a whim of iron’ as others used to call him. And the man who survived countless cigarettes and glasses of red wine, many consumed late into the night as he wrote the week’s Editorials at the last possible moment.

Full tributes to him will appear in next week’s issue (see www.nature.com/jm), but it is appropriate promptly to recall (JM never split an infinitive) some of the highlights of his time at *Nature*. He first took the reins as the editor of *Nature* in 1966. He was the fourth editor — the journal was founded in 1869, and his predecessors had lengthy stints, the first, Norman Lockyer, being in charge for 50 years. John served until 1973, when he was succeeded by David Davies. He then returned in 1980, and I succeeded him in December 1995.

It was during his first stint that he laid the foundations for *Nature* as it is today. Importantly (JM liked to start sentences with adverbs), he threw aside the highly informal and somewhat crony-based system for selecting papers and established a system of peer review. A characteristically readable account of this can be found in his valedictory Essay in his last issue (see *Nature* 378, 521–523; 1995).

This move was not without his own reservations — he liked to say that the 1953 paper on the structure of DNA would never have passed peer review. He never lost his distrust of such refereeing as an obstacle to the truly original, and occasionally dispensed



with it altogether during his first stint as editor.

He also established a strong tradition of journalism in *Nature*. John was a man of many parts but above all he was a journalist, and took pride both in the label and in the craft. He had trained and researched as a physicist, he had an all-consuming intellect, he absorbed research as fast as he could read it — and he was a virtuoso science writer, coming to *Nature* with substantial experience as a newspaper science correspondent. Many leading writers and editors in today’s science media passed through *Nature* during his time, and learned above all how to recognize and seize moments of editorial opportunity even if, many a time, flying by the seat of one’s pants. He established the ‘voice of *Nature*’

in unsigned Editorials (although the voice was often unmistakably his own). And he led the way in developing extensive supplements in which he reported and opined over many pages, often compelling in their narrative, his penetrating perceptions of the state of science and its leadership in this country or that.

So for what else, apart from clouds of cigarette smoke, will John be remembered? Recollections that I have heard from readers over the years include his championing of a research agenda even before many of those in the field had recognized it. Others recall controversial decisions and opinions that were even offensive to some but which, to others more detached from the fray, ‘added to the gaiety of nations’. Many who knew him personally will remember a dry and incisive wit, alongside a strong streak of human kindness.

JM was unique, and those of us who knew him and learned from him will feel the world to be a smaller place in his absence. But his was a powerful spirit, and we continue to thrive on it.

Philip Campbell
Editor-in-Chief, *Nature*

Healthy outlook

China’s first steps towards health care for all will require careful implementation.

On 7 April, the Chinese government formally approved a long-awaited health-care plan. China’s nominally communist regime has, until now, left health care to a wildly profit-driven and generally unreliable system that has cut many citizens off from basic medical attention. The new plan commits 850 billion renminbi (US\$124 billion) over the next three years to begin correcting that situation, and marks the first concrete step towards a goal of providing health care to all Chinese people by the year 2020.

Among the plan’s initiatives are 29,000 new local medical centres and 2,000 new county-level hospitals to reach more rural Chinese; additional training for 1.37 million village-level and 160,000 community-level doctors; a requirement that all doctors spend a year in rural areas; an overhaul of the insurance system; and caps on drug prices. Also, in an effort to make the health-care enterprise more efficient, the plan includes a revamp — or, in many cases, an introduction — of a medical record-keeping system using modern information technology.

Many of these initiatives could have important pay-offs for research into, and control of, infectious diseases. For example, an effective medical-records system could greatly improve the monitoring of emerging diseases such as severe acute respiratory syndrome (SARS) and avian flu, as well as ongoing epidemics

such as HIV and tuberculosis. As things stand now, the data have sizable gaps at the local level. Also helpful will be the Bill & Melinda Gates Foundation's pledge on 1 April to provide US\$33 million to help China's local doctors improve diagnoses of tuberculosis and distinguish between normal and multidrug-resistant strains of the disease.

The government's plan aims to get local physicians up to speed on basic medical care and record keeping. This should make it easier to carry out large-scale clinical trials in rural China, and so help the country realize its potential for translational research. China has perhaps the most diverse disease profile in the world, with huge numbers of patients in various disease categories.

Any plan of such a scale inevitably has its shortcomings and its critics. For example, there is concern that it could stifle drug discovery. Under the plan, the government will produce and distribute medicines deemed to be 'essential', probably based on a list of 300–400 drugs recommended by the World Health Organization. But by depriving some Chinese pharmaceutical companies of income, the regulation could run counter to the government's efforts to stimulate

a nearly moribund new-drug industry. (Defenders of the plan have countered that most truly new drugs won't be included on this list, so that the plan will not discourage innovation.)

Perhaps the greatest missed opportunity is the plan's failure to end the practice of doctors and hospitals adding a 15% fee for drugs that they prescribe. This practice has contributed to widespread over-prescription of drugs, which in turn encourages drug-resistant strains of disease. The plan does call for caps on how much doctors can prescribe for a given illness, and how much a hospital can make from medicines. But that is unlikely to stop them from over-prescribing to meet those limits, or even manipulating figures to expand the limits.

Like every aspect of the new health-care plan, preventing such abuse will require close monitoring at the local level, something that the Chinese leadership has often found difficult. Nonetheless, the initiative has covered an enormous distance in its first step, and will probably be remembered as a landmark in Wen Jiabao's premiership — a period in which China's obsession with all things profitable has given way to a greater concern for the average person. ■

A magnificence to share

Tourism in the Antarctic needs to be regulated, but should not be banned.

Although some will always prefer more cosmopolitan pleasures, there is no denying that, to many, the unspoilt wilderness has a perennial attraction. Shelley captured the appeal well:

I love all waste
And solitary places; where we taste
The pleasure of believing what we see
Is boundless, as we wish our souls to be

Perennial though this desire is, it is also paradoxical. The lure of the wilderness depends to a large part on the absence of humans — and its experience depends on the presence of at least one, and normally more. Few go into the wild alone.

These poetic passions are spurring debate in Baltimore, Maryland, where the signatories of the Antarctic Treaty are celebrating the 50th anniversary of that agreement by holding their first joint meeting with the Arctic Council. The two-week meeting ends on 17 April. Antarctica is the planet's greatest wilderness, and the number of people wanting to visit it increases every year. Growing global affluence — even in these recessionary times — means that more and more can do so. The continent, once the preserve of expeditions whose numbers were counted in dozens, now sees almost 50,000 tourists a season, and the numbers look set to rise.

This poses both practical and ideological problems. Antarctica is a long way from anywhere — that's part of the point — and its waters can be treacherous. Tour vessels get into trouble there with some regularity, and have to be rescued by ships, aircraft and personnel that have been diverted from their mission to support Antarctic

science. Moreover, although the continent is vast, tourists often go to only a few places, thus concentrating their impact on the extremely fragile ecosystem.

For these reasons, the United States has proposed that the Antarctic Treaty be amended to discourage large tour vessels, and to allow no more than 100 people to go ashore at any one time. Such an amendment, which would codify the practices already followed by responsible tour operators — although not by everyone in the business — should indeed be adopted. This cold earth, more so even than Earth in general, needs to be trod on lightly, and there is a compelling need for regulations to ensure that is the case. No one wants to see penguins begging for food like pizza-scavenging racoons in Yosemite.

No one wants to see penguins begging for food like pizza-scavenging racoons in Yosemite.

But the regulations need to focus on impacts — including effects on scientific activities of high value — rather than on total numbers per se. For small and delicate places such as the Galapagos Islands it may make sense to argue, as the Galapagos Conservation Trust does, that every tourist should limit him or herself to a single visit, thus maximizing the number of unique human experiences for a given level of tourism. But Antarctica is far from small. If people want to go there, and they travel responsibly, they should be allowed and even encouraged. The snobbishness that some nature lovers fall prey to — it's for me and my soul-achingly deep appreciation, not hoi polloi — should be resisted. The fact that so many people care so much for natural beauty that they will go literally to the end of the Earth for it is a fine thing; it should be celebrated and indulged as much as is practical.

Purists for whom this will be desecration should start making plans to visit the yet more inhospitable wildernesses of the Moon. If they hurry they may get there before Richard Branson opens a hotel. ■

RESEARCH HIGHLIGHTS

Deafening dolphins

Biol. Lett. doi:10.1098/rsbl.2009.0099 (2009)

Mid-frequency sonar — such as that deployed on military ships — can induce temporary hearing loss in Atlantic bottlenose dolphins (*Tursiops truncatus*), which some people fear can cause them to lose their way and end up stranded on beaches.

Aran Mooney and his colleagues at the University of Hawaii in Kaneohe exposed a trained, captive dolphin to intensive sonar pings to a level of 214 decibels, which induced hearing loss and behavioural changes. Hearing function was measured using an electrode placed atop the head (pictured) to detect auditory evoked potentials in the brainstem. Tests showed that repeated exposures are required to produce the hearing loss.



T. A. MOONEY

NEUROSCIENCE

The thief within

Biol. Psychiatry **65**, 600–606 (2009)

A drug that quiets cravings for alcohol may also soothe the urge to steal in kleptomaniacs.

The drug naltrexone acts by blocking the effects of opioids in the brain. Jon Grant and his colleagues at the University of Minnesota School of Medicine in Minneapolis conducted a double-blind study of the drug's effects on a group of kleptomaniacs who ordinarily steal at least once a week.

Within 8 weeks, the 11 volunteers receiving the drug reported a significant reduction in their compulsion to steal compared with the placebo group. The results are preliminary, but support a link between the symptoms of kleptomania and the body's opioid system.

ASTRONOMY

Twinkle twinkle, lots of stars

Astrophys. J. **695**, 561–573 (2009)

Starbursts, periods of intense star formation in galaxies, have long been thought to be short and frenetic, lasting just several million years or so. But it turns out those episodes were just isolated 'flickers', say Kristen McQuinn of the University of Minnesota in Minneapolis and her colleagues, who measured starbursts in three nearby dwarf galaxies.

They found the flickers to be interconnected parts of longer starbursts,

spread out across each galaxy, and sustained for 200 million–400 million years. These larger, longer starbursts could be responsible for galactic superwinds, which are suspected of being responsible for carrying chemically enriched compounds into intergalactic space.

ANIMAL BEHAVIOUR

Regarding jackdaws

Curr. Biol. doi:10.1016/j.cub.2009.02.062 (2009)

Curious as to whether jackdaws — members of the crow family — could follow human eyes, Auguste von Bayern and Nathan Emery, then at the University of Cambridge, UK, offered food to hand-raised jackdaws. The birds took longer to nab a proffered nibble if the food was the subject of a stranger's stare, and they seemed to be watching the eyes rather than the direction of the head (see variations of head and eye attitudes, pictured below). In a separate experiment, the birds needed moving, rather than static, eye signals from a familiar person to understand communication about the location of hidden food.

The researchers speculate that jackdaws evolved this eye-following ability to interact with one another. However, they add that the birds followed by the study have spent their whole lives with humans.



DNA

Acid-base boogie

Nature Nanotech. doi:10.1038/nnano.2009.83 (2009)

A tiny machine made from DNA can measure the acidity of living cells from the inside.

Yamuna Krishnan and her colleagues at the Tata Institute of Fundamental Research in Bangalore, India, designed three short pieces of DNA that spontaneously assemble into the nanoscale device when they are in close proximity. The device changes shape when in an acidic environment, thanks to changes in how cytosine binds. Fluorescent tags change colour to reflect the shape change — green for neutral, red for acidic. The sensor could be used to study changes in cell pH that accompany viral invasion.

BIOSENSING

Merry-go-round sensing

Optics Express **17**, 6230–6238 (2009)

Detecting nanometre-sized entities in solution is complicated by the fact that they must first slowly diffuse to the sensor.

Stephen Arnold of the Polytechnic Institute of New York University in Brooklyn and his co-workers have found a way to speed things up. By shining a laser on silica microspheres,

they produce 'whispering gallery modes' (WGMs) — light trapped inside the microsphere by reflection that circles endlessly. This causes

ELSEVIER

a short-range electrical field just outside the sphere, which can attract nanoparticles, bringing them to the sensor about 100 times faster than diffusion.

On top of this, the nanoparticles, once trapped, start to orbit the sphere. This produces a shift in the resonant frequency in the WGM, which in turn permits estimation of the size and, by extension, mass of the nanoparticle. WGMs could have applications in biosensing.

NEUROSCIENCE

Tetrapack protein

Cell **137**, 159–171 (2009)

The connections, or ‘synapses’, between neurons change in both strength and shape in response to consistent use — a mechanism thought to underpin learning and memory.

Mariko Kato Hayashi of the Massachusetts Institute of Technology in Cambridge and her colleagues now show that the protein Homer forms tetramers — proteins with four subunits — whose multiple facets could help to coordinate signalling and shape changes at excitatory synapses in the central nervous system.

The researchers show that purified Homer filaments self-assemble with globular hubs of the protein Shank to form a mesh-like polymer matrix that can still incorporate other Homer-binding proteins.

A mutant form of Homer disrupted both the tetramer and the Homer–Shank matrix *in vitro*, and affected synaptic structure and function in cultured neurons and brain slices.

NANOTECHNOLOGY

The new heat order

Nano Lett. doi:10.1021/nl900399b (2009)

The extreme miniaturization that is reached in micro- and nano-electromechanical devices also generates a lot of heat at specific points. This heat has to be dissipated to stop the device from deteriorating, but normal heat-management techniques, such as metal wiring or fluid cooling, do not work well for sources of this size.

Zhiping Xu and Markus Buehler at the Massachusetts Institute of Technology in Cambridge used a theoretical model to evaluate the heat-dissipation performance of hierarchical networks composed of one-dimensional filaments — for example, linked carbon nanotubes. The authors discovered that, even with the same number of dissipating nodes, these structures are much more effective than non-hierarchical configurations.

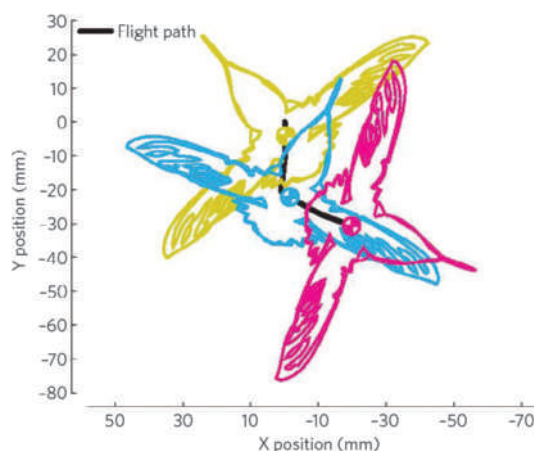
LOCOMOTION

Elegant flappers

Science **324**, 252–255 (2009)

How do birds and bugs that flit through the sky execute precise manoeuvres such as sharp turns and, at the same time, exhibit such stability that crashes seldom occur? Tyson Hedrick of the University of North Carolina at Chapel Hill, and Bo Cheng and Xinyan Deng at the University of Delaware in Newark explain this aerial prowess with a model of ‘flapping counter-torque’.

Focusing on low-speed side-to-side turns of 60° or more (see picture, below), the researchers found that flapping fliers of all sizes — from fruitflies to cockatoos — rely on the turning movement itself to create an asymmetry in velocity between the inside and outside wing and end the turn. That is, they don’t have to actively put on the brakes to avoid overturning; their wings do it for them.



AAAS

CANCER

Dual aspect

Genes Dev. **23**, 877–889 (2009)

Acute myeloid leukaemia in humans can be caused by different genetic changes that translate to markedly different responses to standard therapies.

Scott Lowe at Cold Spring Harbor Laboratory in New York and his colleagues created mice that closely resemble two common variants of the disease. The team irradiated the animals to kill off blood-cell precursors and then injected them with blood precursor cells that had been genetically modified to recapitulate one of two types of leukaemia. As in the human forms, one was very responsive to treatment, the other resisted common therapy.

In addition to providing new insight, the mice may make for a good preclinical model for drug screening, the authors say.

JOURNAL CLUB

Daniel Charlebois

University of Ottawa, Canada

A biophysicist ponders the application of hidden metric spaces to genetic networks.

Complex networks can be conceptualized as a collection of points or ‘nodes’ connected by edges that represent their interactions. The structure and logic of these visualized networks allows mathematical modelling to investigate dynamics such as how information propagates through a system. I am particularly interested in gene regulatory networks — ensembles of molecules and interactions that control gene expression — because of their connection to human diseases such as cancer.

Marián Boguñá and his colleagues suggest that real, observable networks are underlain by geometric frames that contain all nodes, influence topology and guide information-routing decisions. They call these underlying frames ‘hidden metric spaces’ (M. Boguñá *et al. Nature Phys.* **5**, 74–80; 2009).

In addition to the distance between nodes in the observable network, one can measure similarity between nodes — which can be determined by, for example, how many neighbours they share — and abstract it as a ‘hidden distance’. Hidden distances are then used to define the hidden metric space that would place similar nodes closer together, increasing the probability that they are connected and interacting in the network.

A major challenge to applying this framework is explicitly identifying the structure of the hidden metric space for complex networks, for which data sets are often noisy or incomplete. For genetic networks, hidden distances could be abstracted from available data such as tissue-expression profiles. Comparing hidden metric spaces constructed from different data types with known genetic interactions would identify which data are best suited to the process. It will be interesting to see what this reveals in terms of the structure and dynamics of genetic networks.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS



Never mind the size,
look at the genes.

Time to sequence the 'red and the dead'

New projects could tackle the genomics of species both critically endangered and already extinct.

On the first weekend in April, a couple of dozen leading molecular biologists, conservationists and museum curators gathered at Pennsylvania State University in University Park to brainstorm about ways of harnessing the power of the latest molecular sequencing techniques to conservation goals.

"The cost of genome sequencing is falling at an extraordinary rate," says workshop co-organizer Stephan Schuster of Penn State University, who was a driving force behind the 2008 sequencing of a woolly-mammoth genome, the first complete genome of an extinct animal. "Now it is possible to entertain sequencing the genomes of other extinct and endangered species, and the benefits could be huge." Referring to the 'Red List' of highly endangered species drawn up by the International Union for Conservation of Nature (IUCN), Schuster suggests that researchers should plan for sequencing "the red and the dead: a suite of carefully chosen endangered and extinct species."

Devin Locke of Washington University's Genome Center in St Louis, Missouri, showed just how great the potential of the new

technology is when he told the workshop that, by the end of the year, he and his colleagues expect to be producing data at more than 500 times the rate they were capable of in 2006, using a suite of Illumina GAI instruments. Although the Illumina platform is not very well suited to sequencing genomes from species that have not been sequenced before, another high-throughput platform — Roche's Titanium 454 — fits that trailblazing role quite nicely; the technology was used on the mammoth and on the Neanderthal genome that is expected to be published later this year.

Conservation biologists want to use this sequencing power to study the extinction process itself. The idea is to sequence samples from the same species spanning several hundred or even thousands of years, using material from existing zoological, botanical and palaeontological collections. This could give entirely new perspectives on the effects of climate change, disease, invasive species and genetic structure.

In the case of mammoths, for example, there is the promise of sequencing ancient DNA from hundreds, if not thousands, of specimens spanning tens of thousands of years, a record that would take in responses to the climate of the last ice age and its ending.

A recent study showed the promise of such techniques by identifying the likely cause of extinction of the Christmas Island rat (*Rattus macleari*) (K. B. Wyatt *et al.* *PLoS ONE* 3, e3602; 2008). DNA lab work has shown that museum specimens collected before the arrival of black rats on the island were all free of trypanosome parasites,

but many of those collected afterwards were not. "By studying historic extinction events, we're hoping to get a feel for what role such pathogens might be playing in modern declines," says Alex Greenwood of Old Dominion University in Norfolk, Virginia, a co-author on the paper.

Molecular data could also act as an objective metric for endangerment, says Schuster. Once it becomes clearer what happens to genetic

"We're hoping to get a feel for what role pathogens might be playing in modern declines."

HUTTON ARCHIVE/GETTY IMAGES

diversity as a species approaches extinction, he says, “you could then take these hard-core sequencing data to political decision-makers and say, ‘this absolutely has to stop’”.

If the case of the Tasmanian devil is anything to go by, though, politicians may not yet be ready to listen to geneticists. Schuster and workshop co-organizer Webb Miller, also of Penn State University, teamed up with Vanessa Hayes of the Children’s Cancer Institute Australia in Sydney in late 2007 to sequence the genome of the Tasmanian devil, a species on the verge of extinction due to a communicable cancer.

The ultimate goal of identifying the genetic basis for resistance to the cancer (which some animals seem to show) is still some way off, but the researchers have already identified 50 gene markers that allow them to describe the genetic make-up of individual devils. As yet, however, there is no sign that the Tasmanian authorities are prepared to incorporate the maintenance of the genetic diversity that these markers can measure into the conservation plan for the species.

An IUCN representative, who did not attend the workshop, was uncertain whether genomic data would be of widespread benefit for conservation. “The study of species’ genomes could prove valuable, but probably only in a very limited number of cases,” says Jean-Christophe Vié, deputy head of the IUCN species programme in Gland, Switzerland. “For example, if it could help us solve the amphibian extinction crisis by explaining what makes so many amphibians vulnerable to the chytrid fungus, that would be a major contribution.”

Curators at the meeting broadly welcomed the “red and dead” idea. “I don’t think there will be any difficulty convincing museum people of having their specimens used for this work,” says Richard Sabin, curator of mammals at the Natural History Museum in London.

But he points out that extracting DNA requires removing a sliver of tissue from the relevant museum specimen — not something curators take lightly — and the record for samples taken for genetic analysis is not good. Between 2000 and 2007, for example, Sabin says his museum granted 70 requests to sample mammal specimens, resulting in 674 tissue samples being taken. Only two gene sequences from all of these samples made it into the public gene database, GenBank. The reasons for such a low hit rate remain unclear, although failure to extract DNA might be one. But “it is critical for them to report back to us, even if they get nothing from a sample”, says Sabin, who has toughened up the museum’s policy to make sure this happens. ■

Henry Nicholls

Bomb spurs research rally

On 7 March **David Jentsch**, a neuroscientist at the University of California, Los Angeles (UCLA), had his car firebombed outside his house by animal extremists. In response, he formed the group UCLA Pro-Test to oppose such extremism and advocate continued research on animals. The group will hold its first rally on 22 April on the university’s campus.

What sort of research do you do?

I study the neuroscience of mental disorders, with a basic scientific lab approach. Schizophrenia is a major study area. We study how genes in normal animals influence brain function. And we study how potential treatments work. Every project I do in an animal is connected with a human clinical question.

How do you use those animals?

I use rodent and vervet monkey models. We do a lot of work relating genetics to brain function, such as memory and attention. We use invasive procedures or infusions of pharmaceutical drugs. We also use genetic mutation knockouts in rodents. In monkeys, we study genetics and naturally occurring functions. We primarily use non-invasive procedures on monkeys — about 90% are behaviour and genetics studies.

What happened in the attack?

It was 4 a.m. on Saturday 7 March. I was awakened by a loud bang; then I heard the car alarm go off. I went to the window and saw my car on fire. I ran outside to try to put it out, using a fire extinguisher and a garden hose. It was impossible. The gas tank had exploded. When the windows started exploding, I got out of there. The fire got into the trees. If this was July in fire season, I don’t want to even think about what would have happened. It would have been an enormous fire with many homes threatened. No one was injured.

Did you immediately suspect extremists?

I admit I didn’t think of anything else but animal extremists. I am very close to other researchers at UCLA who have been attacked in the past. My immediate reaction was confirmed on Monday 9 March when the North American Animal Liberation Press Office posted a communiqué taking responsibility.

Did you suspect the attack was retaliation for recent US arrests of extremists?

It is hard to say. The communiqué explicitly

linked the attack to the arrests, saying: ‘You hit us; we hit you back harder.’ I think they would have come anyway.

How did you come to found UCLA Pro-Test?

As I walked the UCLA hallways, students, staff and colleagues would approach me. I realized this was affecting many people. The institution was the real target. Others besides myself were traumatized. I knew the extremists would have demonstrations on 22 April for World Week for Animals in Laboratories. I got the idea to do our own rally — to share experiences of our enterprise. I contacted Tom Holder, the UK Pro-Test founder. Within an hour, he was on board.

What do you hope to achieve?

The immediate goal is a rally for pro-research and science advocates to demonstrate our common mission to the community at large. We also want to start a process in which we will no longer let them bomb people and be quiet. The people now traumatized are the ones who know they are next. Those people need support. We are creating a university structure to centralize decision-making on this problem.

Do you think you can change the views of extremists?

I don’t think we can. We can have an open and honest discussion with them to explain the place for animal research in society. But they won’t change our mind anymore than we can change theirs.

How would you suggest that others confront extremism?

In the past, we were always responding to events, instead of being proactive. Everyone at the university knows who the likely targets are. Responses came after attacks, when we were in the thick of the problem. Now we are taking advance steps. There are common-sense things any university can do to be ready. I’m now getting calls from other universities about what they should do. ■

Interview by Rex Dalton

NASA ponders 'carbon copy' of crashed mission

Since the Orbiting Carbon Observatory (OCO) crashed into the ocean minutes after its 24 February launch, researchers at NASA and elsewhere have been working on how else they might get the data on atmospheric carbon dioxide levels that the mission was meant to collect.

Within a week of losing the satellite, NASA, which spent US\$278 million and seven years developing OCO, put together a committee

of two dozen climate scientists to weigh up various options. Should they rebuild OCO with existing designs and launch it as quickly as possible? Start a new design that would take longer to develop? Or fund ground and sub-orbital carbon measurements, while working with

existing greenhouse-gas monitoring satellites such as Europe's Envisat and Japan's Greenhouse gases Observing Satellite (GOSAT, also known as IBUKI).

The case against reincarnating OCO is that the spectroscopy it used to measure carbon levels needed reflected sunlight to work, preventing it from making measurements at dawn, dusk and night. Many scientists, including OCO's principal investigator David Crisp, of the Jet Propulsion Laboratory in Pasadena, California, think that probing the atmosphere

with lasers will eventually offer a way to get round-the-clock data and thus see important effects such as those of nocturnal respiration by soil organisms.

But laser-based systems are technologically challenging. In a recent competition to design atmospheric-science satellites the European Space Agency eliminated a laser-based carbon-dioxide-monitoring mission, A-Scope, citing insufficient readiness. The technology for a

similar NASA mission called ASCENDS (Active Sensing of CO₂ Emissions over Nights, Days and Seasons) is still in development.

Crisp, who chaired the post-OCO working group, says that something like ASCENDS wouldn't launch until at least

2015. Other options — putting OCO-type instruments on weather satellites or on the International Space Station (ISS) — would also take a long time, and in the case of the ISS would miss the polar regions.

So — as expected — the bottom line of the report by Crisp's committee, submitted to NASA on 2 April, was that the agency should build an OCO "Carbon Copy" with the same design and instruments and launch it as soon as possible. Getting the data quickly is "critical to support national policy initiatives", says

"In all our opinions, the need for these data is just as high, if not higher, now as when the observatory was first planned."



Crisp. The repeated mission would cost more or less what OCO cost, and could be ready for launch in the autumn of 2011.

Michael Freilich, head of NASA's Earth-science division, has sent the white paper out for review and says he will make a decision "possibly in May". But even if he plumps for a rebuild, paying for it might be difficult. As

NASA/JPL

Collision debris increases risk to Earth-observing satellites

The collision of two communications satellites on 10 February has significantly increased the risk to Europe's Earth-observing programme.

The European Space Agency's ERS-2 and Envisat missions are 30% more likely to face a catastrophic impact from space debris in the wake of the collision, according to Heiner Klinkrad, head of ESA's Space Debris Office in Darmstadt, Germany. The absolute risk remains small, but there were seven 'near misses' last year in which objects passed within 200 metres of the satellites. The satellites provide

a range of environmental data, including in the case of Envisat some measurements of carbon-dioxide levels similar to, although less precise than, those that were expected from NASA's lost Orbiting Carbon Observatory (see "NASA ponders 'carbon copy' of crashed mission", above).

The increased hazard is the outcome of a collision between a spacecraft in the Iridium satellite constellation and a defunct Russian military satellite (see *Nature* **457**, 940; 2009).

At present, the debris cloud from the collision contains roughly 800 items of 10 centimetres or larger.

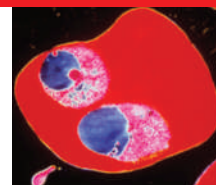
Models show that number could rise to more than a thousand by the end of the month, says Richard Crowther, head of the United Kingdom's delegation to the United Nations Committee on the Peaceful Uses of Outer Space.

The debris field is smaller and more concentrated than many had originally feared, according to Brian Weeden, a technical consultant with the Secure World Foundation, based in Superior, Colorado. That is probably because the two satellites dealt each other only a glancing blow. "It wasn't a dead-on collision," he says.

The concentrated debris field

means that only satellites at fairly similar altitudes face a significantly enhanced risk. That could include some US and Canadian Earth-observing satellites. But spacecraft farther away, such as the Hubble Space Telescope and, when it services Hubble, the space shuttle *Atlantis*, seem relatively safe. US Space Command is tracking the larger pieces of debris, and ESA is making its own radar measurements, with the aim of arranging evasive action, if necessary, to avoid another accident.

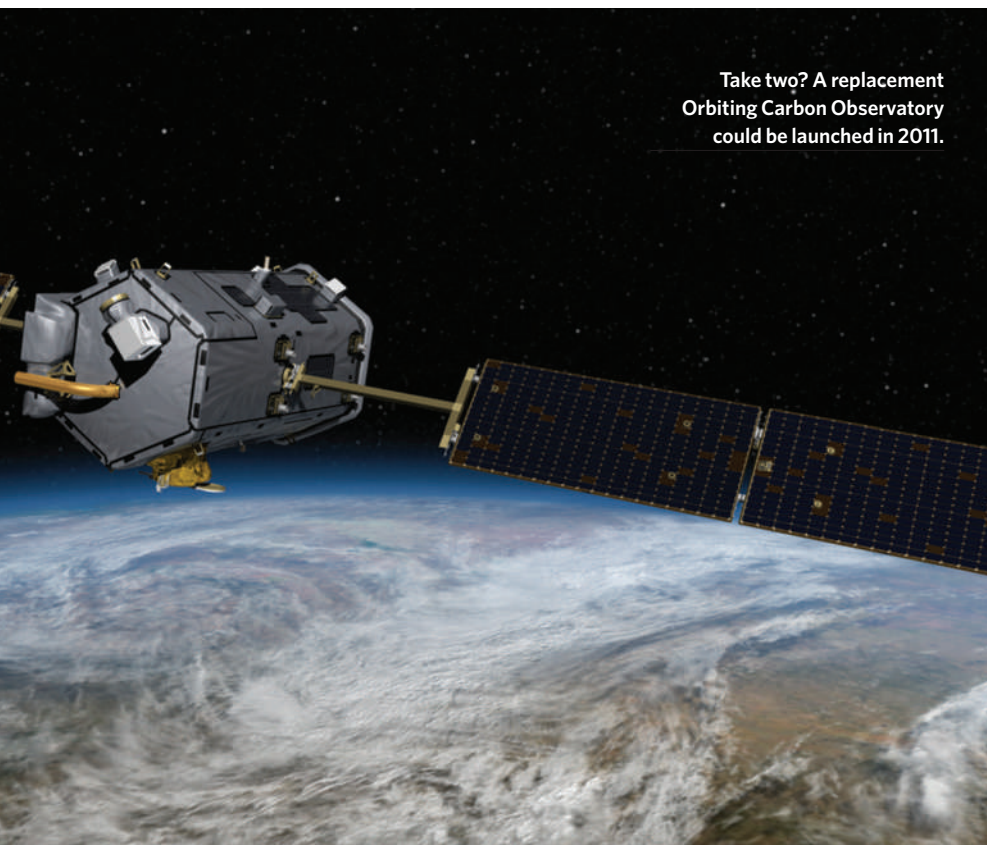
Geoff Brumfiel

**FIGHTING MALARIA**

Drug helps quinine kill resistant parasites.

www.nature.com/news

CNR/SPL



Take two? A replacement
Orbiting Carbon Observatory
could be launched in 2011.

Little progress seen at climate talks

On 29 March, the initial day of the first of three preparatory meetings for the Copenhagen climate-change summit in December, Todd Stern, the US special envoy on climate change, drew thunderous applause when he told the delegates that he was “determined to make up for lost time”. When the meeting ended on 8 April many of the 2,000 delegates had the impression that there was still a lot of making up to do.

Stern told the 175 national delegations that the implementation of a national cap and trade programme could help cut US emissions by around 15% from current levels by 2020, and by 80% by 2050. But he declined to say how feasible such a policy might be politically. Illustrating the challenges, 89 US senators recently voted to adopt a non-binding budget amendment stipulating that climate legislation should not increase gasoline or electricity prices.

Developing countries, including South Africa, India and China, told the Bonn meeting that they expect rich nations to commit to a 40% cut by 2020. Yvo de Boer, the executive secretary of the United Nations Framework Convention on Climate Change (UNFCCC), acknowledged that getting rich nations to agree to 25–40% emissions cuts by 2020 will be “very difficult”.

Another point of contention was the money that poorer countries will need for adaptation programmes. The UNFCCC's Least Developed Countries Fund allows rich countries to support such programmes in the poorest nations and thus meet some of that need, but to date it stands at only US\$172 million. Overall, development agencies talk of a need for sums at least 100 times greater. “There is still no clarity over the scale of financial and technological resources that would be available to developing countries,” said Shyam Saran, the Indian prime minister's special envoy on climate change.

Oil-exporting nations, led by Saudi Arabia, are also interested in a transfer of funds, saying they will demand compensation if a climate agreement cuts oil revenue. “We share the concern for climate change but at the same time we don't want to be a victim,” Mohammad Al Sabban, an adviser to Saudi Arabia's Ministry for Petroleum and Mineral Resources, told reporters in Bonn.

Quirin Schiermeier

OCO was lost just six weeks ago, the only money in the budget that might naturally flow to Carbon Copy is the \$23 million earmarked for OCO's operating costs. That might be a start — and might hold Crisp's team together after its current budget runs out in June — but if Carbon Copy is to fly it needs either new or diverted money.

It might seem that the \$150 million recently added to NASA's fiscal-year 2009 budget for Earth science by Congress, or the \$400 million more given to the same end in the stimulus package, would fit the bill. But there is a queue for that windfall. Landsat, a US Geological Survey land-mapping mission that NASA is procuring, and Glory, a mission due to be launched later this year to study aerosols and clouds, could both use more money if they are to stay on schedule. And other new Earth-science missions recommended as priorities by the National Academies also need to get started. “We have far more to do than the available resources given us,” says Freilich.

There are other options. Carbon dioxide can be monitored from the ground. Pieter Tans, who heads ground-based carbon-cycle monitoring for the National Oceanic and Atmospheric Administration from the University of Colorado in Boulder, points

out that with his \$5-million annual budget, he can monitor 84 spots, mostly in North America, via ground-based sites, aircraft, or ships. That represents approximately half of the world's non-satellite effort to monitor carbon dioxide.

Tans was on the OCO science team, but says he has long worried that excitement about satellites — which are, after all, NASA's stock in trade — leads politicians and policy-makers to neglect ground- and aircraft-based measurements. A more even split in spending between ground and space would allow him to boost his network of sensors by an order of magnitude, he says.

Crisp acknowledges the importance of the ground-based network — OCO needed it to calibrate its indirect measurements — but says there are things only a satellite can do, especially considering the importance of enforcing international climate treaties. “Try putting a CO₂ station the middle of China,” says Crisp. “Try it in the Congo.”

What all concerned agree on is the need to do something soon. Says Ken Jucks, OCO programme manager at NASA, “In all our opinions, the need for these data is just as high, if not higher now [as when OCO was first planned].”

Eric Hand

Forensic labs warn of deuterated drug threat

Recent interest in making drugs in which some of the hydrogen atoms are replaced with deuterium has caused alarm in a key medical speciality that already uses such compounds: forensic toxicology.

Some pharmaceutical companies hope that deuterated drugs will survive for longer in the body, have fewer side effects and combine better with other drugs (see *Nature* 458, 269; 2009). But for the researchers who look for pharmaceuticals in post-mortem examinations or accident investigations the idea is “horrifying”, says Sarah Kerrigan, director of the forensic science programme at Sam Houston State University in Huntsville, Texas.

The problem is that toxicologists already use deuterated versions of pharmaceuticals as reference standards when using gas chromatography mass spectrometry. The reference for a drug of interest usually has three (or more) hydrogen atoms replaced by deuterium, providing a precise signal in the resulting spectrum close to that of the drug being looked for. If the drug of interest were itself also deuterated both compounds would be in the same place. “The bottom line is that we will miss them,” says Kerrigan.

Concert Pharmaceuticals, based in Lexington, Massachusetts, is developing deuterated versions of an HIV protease inhibitor and of the antidepressant paroxetine. “In principle I don’t see that this should cause any problems,” says Roger Tung, chief executive of Concert. He points out that the drugs that Concert hopes to market will have a known number of deuterium atoms, making them easily distinguished from a reference standard with a different number. But Aldo Poletti, from the legal and occupational medicine department at the University of Verona in Italy, says that this overstates the sensitivity of the systems in use.

A reference standard with three deuteriums would in practice be impossible to distinguish from compounds with two to four deuteriums swapped, he says.

Another suggestion of Tung’s — that the unaltered drug could be used as a reference against a deuterated version — also fails to allay the toxicologists’ worries. “If all of a sudden deuterated paroxetine came onto the market and every vestige of the original went off the market and out of everyone’s medicine cabinet, [Tung would be] right,” says Graham Jones, chief toxicologist for the Office of the Chief Medical Examiner in Edmonton, Canada, who maintains a widely used mass-spectrometry database for forensic toxicologists. But with deuterated and undeuterated versions of the drug on the market, there would be plenty of scope for error. “Using the drug as the internal standard would be complete forensic suicide,” says Kerrigan.

“It’s not as simple as using a different drug, we’d have to develop specific methodology,” says Bruce Goldberger, director of toxicology at the University of Florida College of Medicine in Gainesville and editor of the *Journal of Analytical Toxicology*. This is a costly and lengthy process, he says. His lab spends up to \$10,000 for each new methodology validation, which toxicologists must perform every time an aspect of their testing is changed.

Reference standards are bought from specialist companies, such as Cerilliant, in Round Rock, Texas, that provide certified materials. Cerilliant has a number of standards on its books and also synthesizes deuterated versions of known drugs to order. This, though, could lead to new problems if deuterated drugs are patented, and thus need to be made under licence. “We are concerned that the granting of these patents could severely restrict the production of reference materials,” says Mitzi Rettinger, Cerilliant’s vice-president of sales and marketing.

Kerrigan hopes to alert the US Food and Drug Administration to her concerns, and let her colleagues internationally know of the situation through a letter in the *Journal of Analytical Toxicology*. She thinks it will be possible to cope with small numbers of such drugs. “If it is forced on us, we’ll adapt,” says Goldberger. “I’m all for forward-looking medicine,” Kerrigan says, “but the benefits have to outweigh this huge disadvantage.” ■

Katharine Sanderson



It's that damn deuterium again, Horatio.

R. P. JAFFE/CBS PHOTO ARCHIVE/GETTY IMAGES

A synthetic-biology reality check

Is the abrupt closure of prominent player Codon Devices an omen for the field?

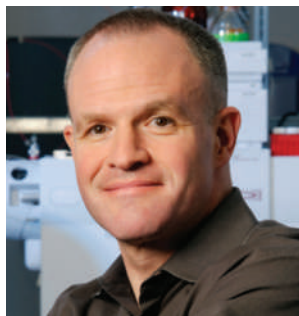
Launched with some fanfare in 2004, Codon Devices of Cambridge, Massachusetts, was touted as a flagship for the industrial applications of synthetic biology, a nascent discipline that applies engineering approaches to the molecular design of biological systems. In late March, however, the company closed its doors.

"I'm personally disappointed," says Codon co-founder George Church of Harvard University, who works on gene sequencing and synthesis technology. "It seemed like an opportunity to bring a lot of money into a field that had a lot of promise and I don't think that's hype; I just think that's a matter of time."

Codon Devices was backed by a range of top-drawer venture capitalists. The scientific founders, alongside Church, were Drew Endy, then of the Massachusetts Institute of Technology (MIT) in Cambridge, who was working on the development of small, reusable genetic components known as BioBricks; Jay Keasling of the University of California, Berkeley, who engineers metabolic pathways; and Joseph Jacobson of MIT's Media Lab, who works on creating molecular machinery. The company started out with a two-pronged strategy: to provide synthetic genes and reagents to order, and to partner with other firms to develop synthetic-biology applications.

But having two different balls in play did not work well. In the first of its two target markets there were incumbent competitors — notably DNA 2.0 in Menlo Park, California, Blue Heron Biotechnology of Bothell, Washington, and Gencart, based in Regensburg, Germany — for whom selling synthetic DNA was the core business. "They tried to do two really hard things in the same small company — to be a successful manufacturing and service company on the one hand, and to be a high-powered R&D company," says John Mulligan, Blue Heron's chief scientific officer. "It's super-hard to combine in one small company those two pretty disparate corporate cultures."

Robert Carlson, a principal at Biodesic — an engineering and design company in Seattle, Washington — and a former colleague of



The fab four: clockwise, from top left, Drew Endy, Jay Keasling, George Church and Joseph Jacobson.

Endy's, points to another problem with combining the two approaches: fears about confidentiality. "Inevitably, when Codon partnered with, say, a company pursuing biofuels, any other biofuels company was bound to feel uneasy about shipping designs for genes or genetic circuits off to Codon."

Codon seemed to recognize the problems in its approach. In March 2008, it received a cash infusion of \$31 million from its existing investors, including such well-connected figures as Vinod Khosla of Khosla Ventures in Menlo Park and Michael Hunkapiller of Alloy Ventures in Palo Alto, California. Three months later, it said it would streamline its business and refocus on applications rather than synthesis. The move may have come too late, says Church. "A lot of our burn rate would be having a sales staff," he explains. "Codon should have stuck with its long-term plan": to provide the "foundational platform" for synthetic biology, as Intel provides a platform for computing.

Still, some analysts are perplexed that Codon's board hasn't explained why the company decided to pull the plug so suddenly. "The lack of explanation as to why Codon

Devices is reported to be 'quietly shutting down' strikes me as very odd," says Steve Aldrich, president of Bio Economic Research Associates in Cambridge, Massachusetts. "I would have thought that the controlling venture investors would be anxious to reassure public markets, which might interpret the abandonment of such a high-profile start-up as a negative judgement on the idea that we would be able to rapidly design, engineer and commercialize biological applications anytime soon."

Drug companies are keen to use synthetic biology to improve drug development, according to Richard Kitney of Imperial College London's centre for synthetic biology, and this raises the question of why the company didn't find a larger concern to buy it up.

Khosla, Hunkapiller, Endy and Jacobson did not respond to requests to comment for this piece. But other scientists say that synthetic biology has

a bright future that need not hinge on the fate of a single firm. "This is a young, exciting and dynamic field, and this is one of the companies that tried to make it and failed in a pretty tough environment," says Sven Panke of the Swiss Federal Institute of Technology in Zurich. "It will take time; that's what you would expect."

Indeed, one of the most promising applications of synthetic biology is thought to lie in the development of better biofuels. In 2007, Church launched a separate synthetic-biology biofuels company, LS9, in South San Francisco, with backing from Khosla, among others. Khosla is also an investor in an earlier Keasling start-up, Amyris Biotechnologies of Emeryville, California, which has begun working on biofuels and is developing cheaper antimalarial drugs in partnership with the Paris-based pharmaceutical company Sanofi-Aventis and the San Francisco-based non-profit Institute for OneWorld Health.

These and other competing ventures may have made life more difficult for Codon, which was recruiting biofuels partners. "You've got the combination of many more [start-ups] getting into biofuels and many more players like BP and Shell," says Kitney. "That must have put quite a lot of pressure on Codon Devices."

Erika Check Hayden and Heidi Ledford

S. OGDEN/SPL

V. STEGER/SPL

V. STEGER/SPL

S. OGDEN

Hope for new telescopes as Mauna Kea plan approved

Hawaiian officials have conditionally approved a comprehensive management plan for the volcanic peak of Mauna Kea, which researchers hope will pave the way for the construction of more telescopes atop the 4,200-metre summit. The University of Hawaii has a long-term lease on a 46-square-kilometre preserve, of which 2 square kilometres are used by 13 observatories.

Some native groups believe that the terrain is sacred and oppose its development. In 2006, a judge reversed a permit for additions to the twin 10-metre Keck telescopes, saying that a summit-wide plan was needed.

The management plan, approved on 9 April by the state's Board of Land and Natural Resources, outlines the role of a management authority set up by the university. Some telescopes previously proposed for the site have been cancelled, but two major projects — the Thirty Meter Telescope and Pan-STARRS, the Panoramic Survey Telescope and Rapid Response System — are pushing ahead.

AMERICA'S SCIENCE ADVISER SPEAKS

John Holdren, US President Barack Obama's science adviser, talks to *Nature*.

On access to Obama:

"I don't have to ask anybody's permission to see the president, except the president."

On cap-and-trade costs:

"The actual numbers are far from devastating. If we had a price of \$100 a ton on carbon, the effect on a gallon of gasoline would be about 30 cents. We all know that the price of gasoline goes up and down by more than that in a week."

On climate legislation:

"If we go to Copenhagen without a climate policy in place, the freedom of our negotiators to negotiate anything meaningful is very limited."

On nuclear proliferation:

"I think we ultimately ought to put all uranium enrichment and fuel reprocessing, if any is done, under multinational control."

Read the full interview at <http://tinyurl.com/dk9w69>.

Capture of wild jaguar raises eyebrows

A federal investigation is scrutinizing the capture and death of the only identified wild jaguar in the United States.

The animal, known as Macho B (right), was caught on 18 February by a biologist team from the Arizona Game and Fish Department and the private Borderlands Jaguar Detection Project (BJDP).

Released with a radio collar, the jaguar became ill, and was then recaptured and killed on 2 March.

The biologists claim that the capture was inadvertent. But an inquiry by the US Fish and Wildlife Service, which monitors such endangered species, began last week after team tracker Janay Brun disclosed that BJDP biologist Emil McCain had directed her to bait a snare with scat from a female jaguar, suggesting that the jaguar was specifically targeted.



ARIZONA GAME & FISH DEPT/AP

Genome Canada cancels stem-cell project funding

Genome Canada, a not-for-profit organization, has pulled its support for an international stem-cell consortium.

The International Regulome Consortium, which involves 12 countries and aims to understand the regulatory networks that guide cell behaviour, expected Genome Canada to provide Can\$20 million (US\$16 million) over 5 years towards the Can\$80-million project.

Genome Canada's head Martin Godbout says that the organization decided not to continue its support after an interim review of the project's science, management and budget recommended substantial changes. The consortium head, Michael Rudnicki, says that the decision was made because the organization lacked the funds after receiving no money in Canada's 2009 federal budget (see *Nature* 457, 646; 2009).

"This is about the conservative government failing to support science," he says. Rudnicki says he is working to organize funding and revamp the structure so that the consortium can continue.

Argentina's dengue-fever outbreak reaches capital

A resurgence of dengue fever in Latin America has hit Buenos Aires. On 9 April the city confirmed that the number of cases had exceeded 150.

Argentina is in the middle of its worst dengue epidemic since records began almost a century ago. Nationwide, more than 10,000 cases have been confirmed and specialists estimate a true total of about 30,000. The country is also dealing with its first reported cases of the potentially lethal haemorrhagic form of the disease. Neighbouring Bolivia has been hit by as

many as 114,000 dengue cases this year.

Dengue has become the world's most widely spread vector-borne disease over the past decade, according to Ricardo Gürtler, a dengue researcher at the University of Buenos Aires. Largely driven out of Latin America in the 1950s and 1960s, dengue's comeback has been linked to factors such as climate change, urbanization — which has been particularly rapid in Latin America — and decreased use of pesticides that reliably kill the mosquito vector.

Japanese stimulus provides green boost

Japan's Prime Minister Taro Aso has announced a ¥15.4-trillion (US\$154-billion) stimulus package that includes measures to spur on the country's green-technology industry.

The package, announced on 10 April, allocates ¥1.6 trillion to low-carbon technologies, including plans to equip schools and homes with solar panels to increase solar-energy capacity 20-fold to 280 gigawatts by 2020. Subsidies in the range of ¥250,000 for electric and hybrid cars aim to create sales of 1 million environmentally friendly cars by the same deadline.

Aso is pitching the project — which still needs parliamentary approval — as a bridge to a future in which Japan is less dependent on cars and conventional electronics manufacturing and thrives on environmental technologies.

A supplementary budget including the green stimulus will be submitted to parliament on 27 April.

Correction

In 'Open-access policy flourishes at NIH' (*Nature* 458, 690; 2009), the number of articles downloaded from PubMed Central represent the average usage on a typical weekday, not a monthly total as we stated.



THE TEST OF INFLATION

As the launch of the Planck spacecraft approaches, Eric Hand investigates what the mission could mean for the predominant theory of the moments after the Big Bang.

Space is cold. But Planck will be even colder. At the heart of the European Space Agency spacecraft, chilled to 0.1 kelvin by the most sophisticated cryogenic system ever put into space, a collection of gossamer threads will be suspended in near vacuum. These threads, looking like an array of spiderwebs, will gather photons from the cosmic microwave background (CMB) — the afterglow of the Universe's creation.

The extreme cold will be necessary to fulfil Planck's mission to make ultraprecise maps of the CMB. Primordial photons, the oldest light in the Universe, stream through every cubic centimetre of empty space at an average temperature of just 2.7 kelvin. The temperature of the photons will vary very slightly, depending on which part of the sky they are coming from (see 'Mapping the cosmos', opposite).

Planck's supercold detectors are designed to measure these differences in temperature at the level of less than a millionth of a kelvin (see 'Cooling Planck down', page 823). That extraordinary precision, say the spacecraft's designers, means that once launched — a milestone now scheduled for 6 May — the probe could over its 2-year lifetime settle questions that have roiled the astrophysical community for a generation. "We have to dig deeper and Planck is our next best chance to do that," says project scientist Jan Tauber, based in Noordwijk, the Netherlands.

For nearly 30 years, says Tauber, the thinking of cosmologists has been guided by a theory called inflation, which tries to explain how the Universe evolved in the moments after the Big Bang. Inflation has passed every observational test to date, mainly by predicting the statistics of the temperature variations in the CMB seen by Planck's

predecessors. But Planck's extreme sensitivity to the variations will put inflation to its most stringent test yet — and will either vindicate it, or demolish it in favour of some rival theory, of which there are several.

Indeed, inflation is so important to modern cosmology that the ultracool, 2-tonne, €600-million (US\$800 million) spacecraft is in a very hot competition with dozens of ground-based and balloon-borne experiments, all pursuing the same goal: exquisite measurements of the CMB (see 'The race for B-modes', page 824). "This is a very big race," says Michael Turner, a cosmologist at the University of Chicago in Illinois. "This is Swedish gold."

A helping hand

The Big Bang would have stayed pretty small without some sort of boost. To reconcile quantum theory with cosmology, physicists would like to believe that the primordial Universe started out just 10^{-35} metres across. That leads to a contradiction. The age of the Universe is reasonable well known: 13.7 billion years. If it started off as small as the theorists would like and expanded only at the rate that now prevails, it would still be able to fit comfortably within the full stop at the end of this sentence.

Inflation explains this apparent paradox by postulating a spectacular expansion in the very first moments (see 'Timeline of the inflationary Universe', page 822). There is no easy analogy for how furious and quick an expansion it was, but as an example, in one simple inflation model, that same infinitesimal Universe could have spread to something like $10^{1,000,000,000,000}$ metres across. That's a one followed by a trillion zeros. And it would have done so in

S. CORVAJA/ESA

a trillionth of a trillionth of a trillionth of a second. Not even light could keep up: the farthest a photon could have travelled since the Big Bang — a scale known as the horizon distance — is ‘only’ about 10^{27} metres. (This is not a contradiction with relativity: no two particles located at the same point during inflation ever had a relative speed greater than light. The explosive speed refers instead to the scale of the Universe as a whole.)

In addition to explaining the Universe’s colossal size, inflation neatly solves many other problems. It explains why the Universe appears geometrically flat rather than curved — think of a balloon that has been blown up so far that its surface looks like an infinite plane. And it explains how a Universe that easily could have looked utterly different in every direction in fact looks pretty much the same, with the same average density of galaxies and the same average CMB temperature.

On top of all that, inflation explains the galaxies of which we are a part. Although it made the infant Universe almost completely flat and uniform, inflation also had to obey the dictates of quantum mechanics, which produced the tiniest of fluctuations in density from point to point. So some parts of the Universe would have ended up denser than others. These denser regions would have become the seeds around which galaxies and stars would gravitationally coalesce. And earlier observatories such as NASA’s Wilkinson Microwave Anisotropy Probe (WMAP), which was launched in 2001, have shown that these fluctuations in the CMB not only exist, but have precisely the kind of size distribution predicted by inflation¹.

Naming the unknown

Yet for all its explanatory power, inflation has its problems. For starters, no one knows what did the inflating. Theorists describe the ‘force’ as a field and give it a name — the inflaton — but the mystery remains. It is the same frustration that bedevils astronomers studying dark energy, an unknown force that accounts for three-quarters of the energy in the Universe and still accelerates its expansion. Could the cause of inflation also be the driver behind dark energy? It is an interesting similarity, but they act on vastly different scales; dark energy is a flea to inflation’s elephant. “It seems unlikely that they’re related,” says Turner. “Which is a good reason to pursue that idea,” he adds impishly.

A bigger problem for inflation, according to Paul Steinhardt, a physicist at Princeton University in New Jersey, is not so much what it is, but how it stopped. “Once it starts it never ends,” says Steinhardt, who was one of inflation’s founding fathers in the 1980s, but is now one of its chief critics. There is no obvious reason why the ultrarapid expansion should ever slow down to the much more modest rates of expansion seen today. “At first,” says Steinhardt sarcastically, “this was celebrated.”

The ‘celebration’ of this notion of inflation without brakes is a jab at Steinhardt’s colleague and sometimes rival Andrei Linde: a physicist at Stanford University in California. (Linde, Steinhardt and Steinhardt’s graduate student of the early 1980s often share in the credit given to Alan Guth, now at the Massachusetts Institute of Technology in Cambridge, for originating the theory that Guth first set out in 1980 (ref. 2).)

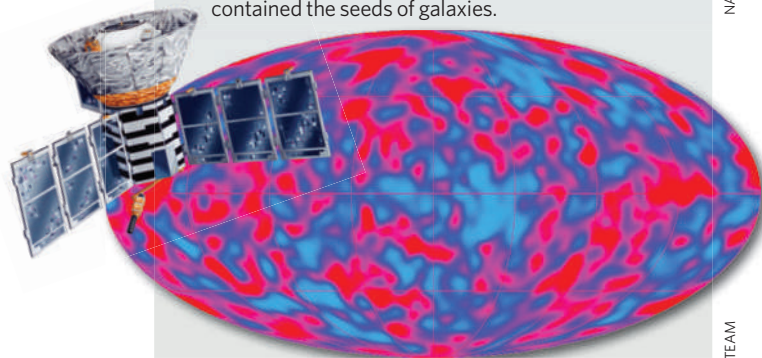
By 1986, Linde theorized that, because of quantum fluctuations, some portions of the Universe might feel the inflationary force more strongly than the average portion³. The result in each case would be a localized bulge. Because of inflation’s immense strength, however, the bulge would quickly take over and inflate into a whole new universe,

Mapping the cosmos

When the cosmic microwave background (CMB) was first detected in the 1960s, it was just a uniform hum on the sky. But a series of spacecraft, built for precision cosmology, has revealed the CMB in ever-increasing detail. “Last year’s discovery is this year’s calibration,” says Nobel prizewinner George Smoot, who was a principal investigator for COBE. Shown here are three important spacecraft and sample all-sky maps (simulated for Planck).

Cosmic Background Explorer (COBE): 1989 launch

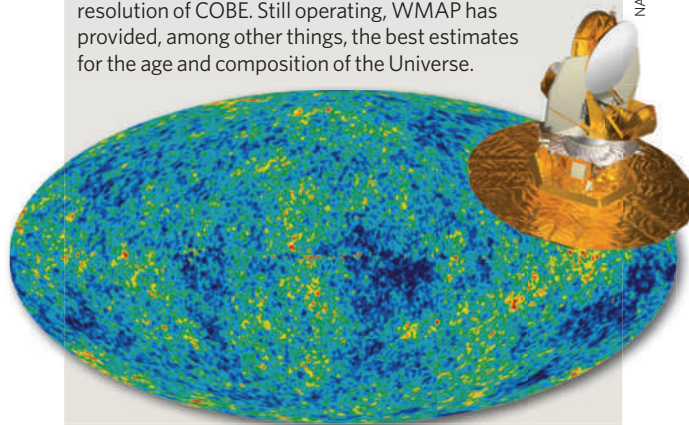
COBE saw temperature differences at large angular scales by comparing big patches of sky. This anisotropy — at levels of 1 part in 100,000 or about 30 microkelvin — contained the seeds of galaxies.



NASA

Wilkinson Microwave Anisotropy Probe (WMAP): 2001 launch

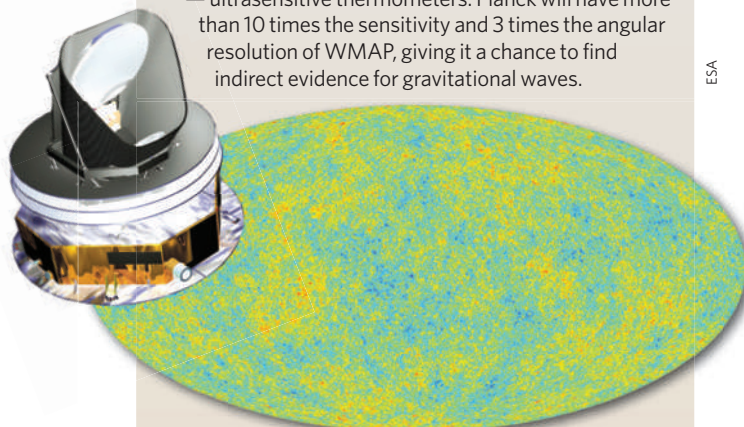
WMAP maps at much smaller angular scales. It has 45 times the sensitivity and 33 times the angular resolution of COBE. Still operating, WMAP has provided, among other things, the best estimates for the age and composition of the Universe.



NASA/WMAP SCIENCE TEAM

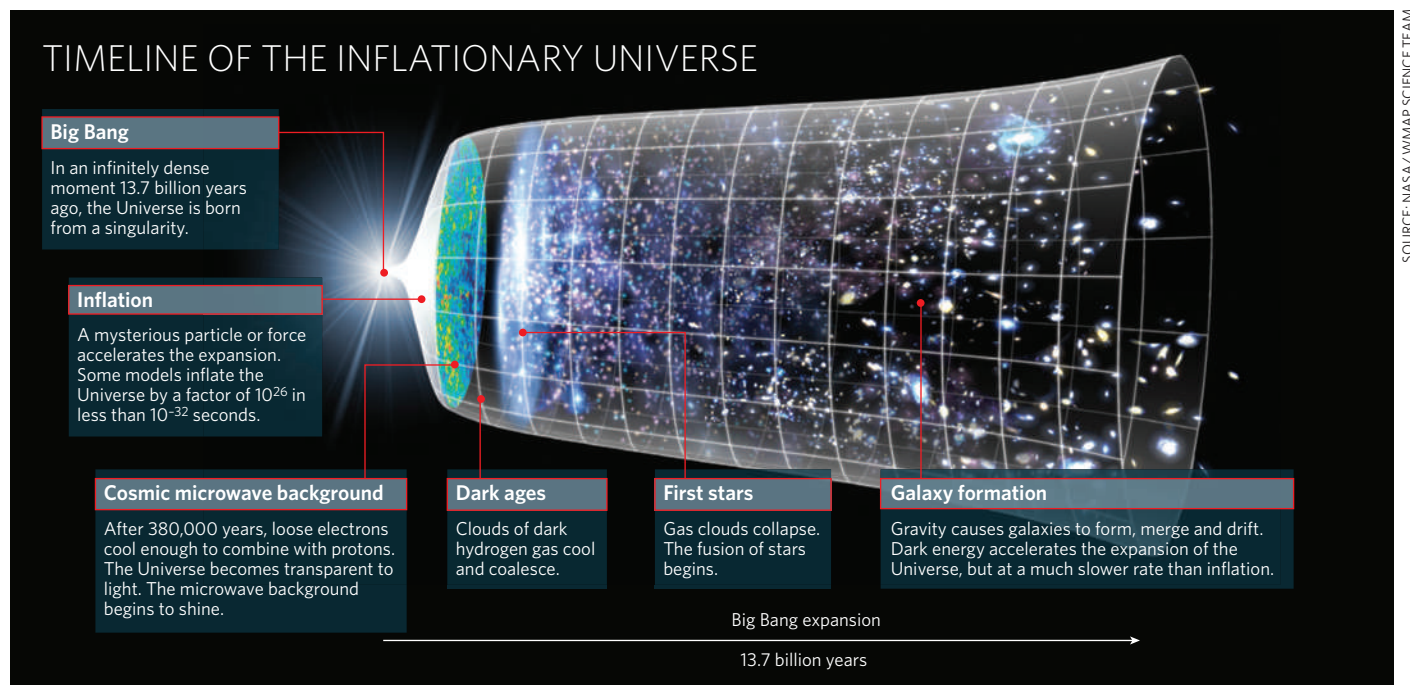
Planck: 2009 launch

Planck will be the first CMB spacecraft to carry bolometers — ultrasensitive thermometers. Planck will have more than 10 times the sensitivity and 3 times the angular resolution of WMAP, giving it a chance to find indirect evidence for gravitational waves.



ESA

TIMELINE OF THE INFLATIONARY UNIVERSE



which would be attached to the old one by nothing more than a quantum-scale thread. To residents of the old universe, that connection would be too small to see; they would never know that a new creation had happened right in front of them. To residents of the new universe, it would look like a new Big Bang of their own. Indeed, universe could sprout on universe, like a fractal. The multiverse, as Linde calls it, would go on forever.

But where Linde saw something creative, an endlessly budding tree, Steinhardt saw something more like an aneurysm, debilitating if not deadly to the overarching theory. If the new offshoot universes tend to take over, the ones left behind become rare islands. What you think is going to be a typical universe is quickly left behind by newer, emerging universes. Conceivably, moreover, the physical laws could change with each emergence, so theorists couldn't know if the physics seen in our Universe governs overall. "The part that we observe — our Universe — exists almost nowhere. It's the rare region," Steinhardt says.

There is a third issue with inflation — less a problem of the theory itself than an issue of its reach or scope. As it stands, inflation is not a theory of how it all began, but a theory of how it all began just after the beginning. Call it a morning-after theory.

In inflation, the Big Bang itself remains an unknowable, infinitely dense moment of time and space called a singularity. "The biggest weakness is the notion that we can get by in a theory of cosmology without understanding the singularity," says Neil Turok, the director of the Perimeter Institute for Theoretical Physics in Waterloo, Canada. "It's like, 'let's just start the clock a little bit afterwards'."

Steinhardt points out that the term Big Bang was originally coined by Fred Hoyle, a staunch opponent of the idea, as a way of mocking the notion of a cosmos suddenly appearing in the clap of a magician's hands. Ironically, the term stuck. "Some people like the idea of there being a moment of creation," Steinhardt says.

But he isn't one of them. Nor is Turok. In 2001, the two physicists proposed a radical alternative to inflation called ekpyrosis, from the Greek for 'out of fire'⁴. It grew out of discussions with string theorists, who see the visible world as inhabiting lower-dimensional membranes,

or branes, in a universe made up of at least 10 dimensions. Steinhardt and Turok proposed two universes on separate three-dimensional branes that would oscillate back and forth along a mutually perpendicular dimension, like sheets hung out to dry on parallel washing lines. Every trillion years or so, after each universe had dissipated into darkness during an expansive phase, the two branes would approach one another and collide, releasing a fireball of energy to start each universe afresh. "It would mean that the Big Bang wouldn't be a beginning but a collision," says Steinhardt.

Avoiding the singularity

Ekpyrosis mimics many of inflation's appealing features, with some key differences. Almost by definition, it avoids the singularity, because it describes the collision of branes as a continuous process. "We have to resolve stuff from before to after," says Turok. "Inflation can get away with ignoring it." Turok says string theory has given him tools that now allow him to describe, through ekpyrosis, a crunch that avoids a singularity altogether.

Linde says he will believe it when he sees it. He has pointed out mathematical problems in ekpyrosis, and forced its proponents to revise their thinking since their original proposal. He occasionally enlists the help of string theorists, but he says he now has a hard time getting them to pay any attention to inflation's alternatives. If anything, the big advance in the past five years is that string theorists are finally finding linkages between their work and inflation. Ekpyrosis, says Linde, is "like a house of cards".

Most theorists would say that inflation is still the best game in town. But Daniel Baumann, a postdoctoral researcher at Harvard University who has worked with Steinhardt and Turok, and who is the lead author on a recent paper that maps out the theoretical landscape of inflation and its alternatives⁵, speaks for many when he adds that it is also "something that should be challenged".

Generational politics might be playing a role in the debate, says Linde. Young scientists don't want a career based on polishing an existing theory, he says. They would rather strike out on their own with something revolutionary — as he did when he first published on inflation in his early 30s

"There are people who like making a mess, and there are people who like cleaning it up."

— David Spergel

Cooling Planck down

The Planck spacecraft, once it settles in a gravitational dead spot 1.5 million kilometres from Earth, will face away from the Sun. "No sunshine ever falls on the core," says Charles Lawrence, the US project scientist for the mission at the Jet Propulsion Laboratory in Pasadena, California. "No earthshine. No moonshine. It's completely in the dark." And it must stay incredibly cold to have any chance of detecting the tiny temperature differences between photons.

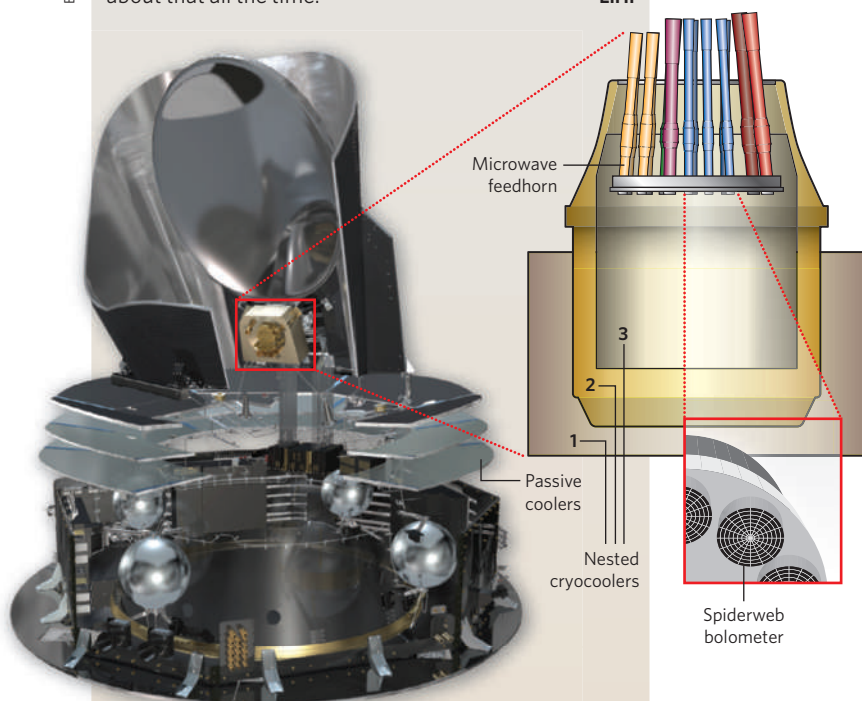
A passive system of three stacked disks starts the elaborate process of dissipating the heat from Planck's posterior, which will face the Sun and reach 380 kelvin. Working like a car radiator, the disks spew heat away, and bring the temperature down to 50 kelvin.

Once Planck reaches orbit, three cryogenic coolers, nested like Russian dolls, will turn on. The first one, using the heat-sucking ability of decompressed hydrogen, drops the temperature to 20 kelvin. The second, a mechanical compressor, sends temperatures plummeting to 4 kelvin. And the final cooler, which uses the different thermal capacity of helium and a helium isotope with one less neutron, brings molecular movement nearly to a standstill: a tenth of a degree above absolute zero.

In this innermost sanctum lie 52 bolometers, which take the energy of electromagnetic radiation — cosmic microwaves in this case — and convert it into heat that is measured by a thermometer. The bolometers are made from fibres of gold-coated silicon nitride, a hundredth the width of a human hair. They hang in place like the filigree web of a spider: mostly empty space to allow cosmic rays, or charged particles, to pass through without confounding measurements. The spindly threads are the right size, however, to react to cosmic microwave background radiation. Corrugated antennas that poke through the nested cryocoolers gather microwaves at different frequencies and funnel them to the bolometers.

It's mind boggling, says Lawrence. Planck will gather photons that have been moving since the Big Bang and show if some are a ten-millionth of a degree hotter than others. "It's not the movie, it's not the historical novel, these are the original photons," says Lawrence. "They've been travelling 13.7 billion years to us. I think about that all the time."

E.H.



— whereas those who erected the framework of inflation have a vested interest in defending it. But that doesn't explain the iconoclasm of older scientists such as Turok, and especially Steinhardt, who helped establish inflation. With them, according to David Spergel, an astrophysicist at Princeton University who is on the WMAP team, it is just a matter of personal style. "There are people who like making a mess," he says, "and there are people who like cleaning it up."

Variations on a theme

There could be another explanation for the restlessness among theorists. They may just be bored with a lack of new data. The WMAP data have nearly been wrung dry. Existing CMB data somewhat constrain the development of new theories, which is especially challenging if you want to invent a radical one like ekpyrosis. But there are few barriers to creating a new flavour of inflation. Variations on the theme are proposed (and often, quickly quashed) all the time: such as racetrack inflation, multi-field inflation and hyperextended inflation. Each of them satisfies the basic tenets of inflation, but the differences — primarily the shape and duration of the inflationary pulse — cannot be distinguished with existing data.

Planck could change that situation. Two tests will be particularly important, not only in weeding out the imposters, but in constraining key parameters such as when inflation began — if it happened at all — and how long it lasted. One of these tests will be for the 'Gaussianity' of the CMB. The mottled temperature variations of the CMB sky seem random, but the likelihood of a hot spot sitting next to a cool spot may follow some rules. If a perfect and simple inflation governed the early Universe, then these shifts from hot to cold — the variations of the variations — should have the Gaussian shape of a bell curve. Ekpyrosis, and many of the more complicated versions of inflation, have characteristic deviations from the Gaussian curve. "The implications of non-Gaussianity are so profound, you really want evidence that it's there," says Spergel.

Last year, using WMAP data, Benjamin Wandelt, a Planck scientist from the University of Illinois at Urbana-Champaign, concluded to his own surprise that the CMB seemed to have significant non-Gaussianity⁶. But the result, based on months of intensive computer time, could easily evaporate even though it's supposed to be 99.5% certain. "It's at the level of being tantalizing but not conclusive," says Wandelt. The WMAP team, which has performed analyses of its own, has so far resisted a declaration of non-Gaussianity. But the number of studies claiming non-Gaussianity continues to grow, including a recent one led by Christoph R  th of the Max Planck Institute for Extraterrestrial Physics in Garching, Germany⁷. "There's a consistent picture of inconsistencies," he says. Without Planck, however, the existing hints are unlikely to congeal.

A bigger test for inflation, the 'smoking gun' that has scientists building microwave telescopes at the South Pole and sending balloons nearly to space, is the search for B-modes. B-modes are a special type of polarization that may be present as markings on the CMB. If they exist, they would be indirect evidence for the gravitational waves that should have accompanied inflation.

Ekpyrosis is clear about this: it predicts no gravitational waves whatsoever, and so B-modes should be absent. With inflation, the B-mode story, like the prospects for Gaussianity, is more complex. The simplest form of inflation should produce a relatively large B-mode signal, just within reach of Planck. But in other forms of inflation, the signal could be tiny. If Planck detects a B-mode signal and

The race for B-modes

Gravitational waves right after the Big Bang may have tattooed the cosmic microwave background (CMB) with signals that would be evidence for inflation. "It would be an almost guaranteed Nobel prize to whatever group was able to detect this," says James Hinderks, a postdoctoral researcher at Goddard Space Flight Center in Greenbelt, Maryland.

New generations of ground-based, balloon-borne and satellite experiments are vying to be the first to claim a detection of the signals, called B-modes, which would be a measure of variations in the way the CMB is polarized across the sky. But B-modes don't definitely exist. And if they do, they are at least a thousand times fainter than the CMB. A simple inflation model would put B-modes within the range of most experiments — but other inflationary theories set a much lower bar.

Hinderks is working on one of the experiments, the Primordial Inflation Polarization Explorer (PIPER) that, if it flies in 2013, would be an order of magnitude more sensitive to B-modes than Planck. But to do that, he'll have to lift a 3,500-litre tub of liquid helium with a stratospheric balloon. A microwave telescope would peer out of the bucket towards the 'southern hole' — a small window in the sky, nearly free of dust from the Milky Way, with a clear shot all the way to the CMB. Ground-based telescopes in Antarctica would do the same thing during long polar winters — when the atmosphere is extraordinarily dry and free of water vapour, which emits microwave radiation. In space, Planck will eliminate the confounding effect of foreground microwaves and look for a B-mode blip at much larger scales on the sky. The Planck team plans to analyse and release its first data set by the end of 2012, by which time some of the other experiments could have results.

There are trade-offs in each approach. Ground-based telescopes can be built cheaply with the latest technology; satellites are expensive and slow to develop, but are able to support clear and wide views of the sky. Balloons occupy a middle ground. Total US research funding for a dozen suborbital CMB experiments is about US\$25 million a year, much less than any space-based successor to Planck, which would cost many hundreds of millions of dollars. "\$25 million is peanuts," says Stephan Meyer, a physicist at the University of Chicago in Illinois, "and the science we get out of this is fabulous."

But regardless of the observation post, all will have to make very precise temperature measurements: the millionths and even billionths of a degree differences between Big Bang photons from one part of the sky to another. "You're talking about measurements of temperature variations at the nano-Kelvin level," says Michael Turner of the University of Chicago in Illinois. "That's how important this [work] is, that people would worry about [doing] this."

E.H.



In Antarctica: BICEP (on the central blue building) looks for B-modes from the ground.

S. RICHTER/BICEP

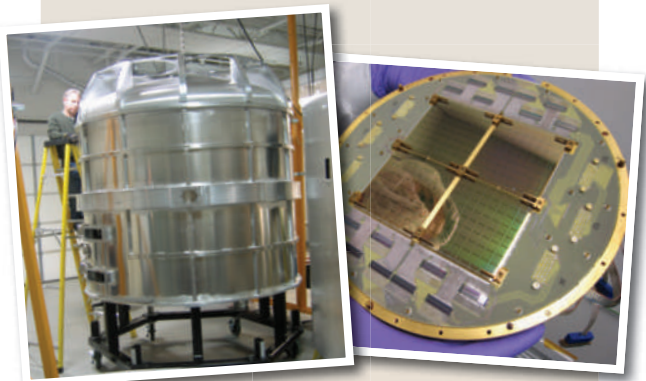
non-Gaussianity, inflation would certainly be "in more of a corner", according to Turok. Unable to resist a poke, he adds: "It's already in a corner."

A non-detection of B-modes by Planck would put the pressure on the ground-based and balloon experiments, some of which claim to have B-mode sensitivities about an order of magnitude better, although they cover far less of the sky. But Andrew Lange, a physicist at the California Institute of Technology in Pasadena, says there are limits to how far they can go. Lange has probably done more than anyone to pioneer the technology of bolometers, high-resolution microwave thermometers that will be used on Planck. If the B-mode signal is too weak, then the contemporary microwave emissions of the Universe (mostly from the Milky Way) will drown out the relic signal no matter how good the bolometer.

At that point, the only opportunity to study inflation would be to observe the gravitational waves directly, via a mission such as the Big Bang Observer. A proposed successor to the Laser Interferometer Space Antenna — another far-off space mission to detect and measure gravitational waves — the Big Bang Observer would have the tough task of directly detecting the low hum of inflation's gravitational waves, rather than picking them up indirectly in the CMB, as the B-mode experiments try to. But the Big Bang Observer's technology, and its funding, are still just theoretical. And without hints of a signal from one of the B-mode experiments, the project will be hard to justify. If the bang was indeed ekpyrotic, there may be no Big Bang gravitational waves to observe, however subtle the instruments.

And so any talk of Swedish gold — Nobel prizes — for the first finder of a B-mode, and probably also for the proposers of inflation, is either spot on or way too premature. Lange, for one, has bets on the former. In the next few years, he will have bolometers of his devising in space, on a balloon and at the South Pole. But he is also steeling himself in case of the latter. When he gives talks about the search for B-modes, he always inserts pictures of people chasing geese. If the current generation of experiments fails to detect anything, he will humbly accept that Nature has put some things beyond his reach, and his lifetime. "Then, personally, I go do something else."

Eric Hand is a reporter for *Nature* based in Washington DC.



The SPIDER balloon experiment will use a cryostat of liquid helium (far left) to cool instruments such as polarization-sensitive focal planes (near left) in its search for B-modes.

1. Wright, E. L. et al. *Astrophys. J.* **396**, L13–L18 (1992).
2. Guth, A. H. *Phys. Rev. D* **23**, 347–356 (1981).
3. Linde, A. D. *Phys. Lett. B* **175**, 395–400 (1986).
4. Khoury, J., Ovrut, B. A., Steinhardt, P. J. & Turok, N. *Phys. Rev. D* **64**, 123522 (2001).
5. Bauman, D. et al. Preprint at <http://arxiv.org/abs/0811.3919> (2009).
6. Hand, E. *Nature* doi:10.1038/news.2008.755 (2008).
7. Räth, C., Morfill, G. E., Rossmannith, G., Bandy, A. J. & Górski, K. M. *Phys. Rev. Lett.* **102**, 131301 (2009).



TWO BY TWO

Could genes explain the remarkable rate of identical twins born in some remote villages around the world? **David Cyranoski** investigates a long-standing biological curiosity.

In December 2008, Bruno Reversade travelled to India in pursuit of some spit. The journey took him first to the north-eastern city of Allahabad, and then a further 10 kilometres to Mohammad Pur Umri, a farming village enclosed by mud walls. There he asked some of the 2,000 or so residents to deposit samples of their saliva in a cup specially designed for the purpose. Fifty-five of them complied.

This is not the first time that these villagers have sacrificed their body fluids for science. Mohammad Pur Umri has become somewhat famous, not for the milk or mustard that provides the villagers with their livelihood, but for its prolific production of identical — monozygotic — twins. Globally, only 1 in every 250 to 300 births are identical twins. In Umri, roughly one in ten is of this type, births that the villagers — including the twin village leaders — call “gifts from God”.

Reversade is looking for genes that might be responsible for this gift. “Every 50 seconds a pair of natural clones is born. It’s more frequent than some of the most frequent genetic diseases,” says Reversade, a developmental biologist at Singapore’s Institute of Medical Biology.

“It can’t be random.” Many scientists disagree, arguing that chance could fully account for this cluster of cases and for every twin birth besides. Aside from genes and chance, theories abound for what causes a fertilized egg to produce a pair of monozygotic twins. None is well accepted. Nor do scientists understand how entities that are apparently genetically identical can come to have such different personalities and disease susceptibilities. Environmental factors explain some, but not all, of these differences.

Over the past few years, scientists have been using new genetic and cell-biology techniques to attack these questions. Reversade is gathering samples from three ‘twin towns’, and using genomic analyses that, he hopes, will point to a common molecular pathway involved in twinning. Embryologists and obstetricians are looking for clues in assisted reproduction (see ‘Making twins’), which is known to promote monozygotic twinning as well as high rates of dizygotic twins, which result from the transfer of multiple embryos. Theories on how identical twins come to differ are also being

overhauled, with some studies even suggesting that genetically identical twins may never, even at the earliest stage of development, have been genetically identical. “This research just hasn’t been part of people’s thinking,” says Judith Hall, a prominent twinning researcher at the University of British Columbia in Vancouver.

Reversade became interested in monozygotic twinning after cutting frog embryos into halves and watching through a microscope as they developed into identical embryos. “Pure awe,” he says. His goal is to understand why cells that are acting together to form an embryo split off and

start building a whole new organism, something they can do early, when the embryo is just a few cells big, or as late as two weeks into development. Conjoined twins can result if the embryo splits too late or incompletely. Reversade says that twinning offers the best way to study ‘regulative development’ — the interaction between cells that informs each one when to follow the pack and when to act alone. He moved to Singapore in February 2008 after landing the government’s first A*STAR Investigatorship, a US\$500,000

“Every 50 seconds a pair of natural clones is born.”

— Bruno Reversade

B. REVERSADE



Bruno Reversade (left) is collecting DNA from various 'twin towns'.

the 22 autosomal chromosomes. But to make the hereditary pattern work, the gene must have 'variable penetrance', such that some women would not bear monozygotic twins even though they carried a copy of the gene. "Variable penetrance is of course a 'black box,'" says Reversade. "Why don't we see more twins?" One reason, he suggests, is that some twins 'vanish' — meaning that at least one of the two dies — during pregnancy.

"Humans do it any old time and can easily miss the right time."

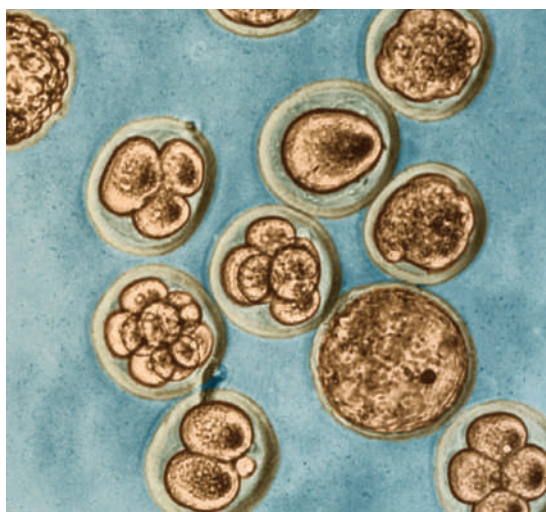
— Judith Hall

per year grant modelled on the Howard Hughes Medical Institute awards, having convinced the funders of the feasibility and importance of finding a twinning gene or genes.

Twinning is often assumed to run in families, and in the case of dizygotic twins, caused by the release and then fertilization of two eggs, scientists agree that this is the case. So far though, researchers have only found genes that are weakly correlated with dizygotic twinning. The picture for monozygotic twins is even less definitive. In large-scale studies, family members of the mothers of monozygotic twins do not seem more likely to have monozygotic twins themselves¹. But some families and some communities do produce identical twins in numbers that seem to defy this interpretation. Reversade is visiting them one by one.

Two years ago he travelled to Jordan to collect saliva samples from members of a family with 15 pairs of monozygotic twins. The family tree fits a pattern in which a dominant gene — one that would cause monozygotic twinning even if only one copy is present — is on one of

According to a widely cited estimate by Charles Boklage, a behavioural and developmental biologist at East Carolina University in Greenville, North Carolina, at least 12% of natural conceptions will produce twin embryos. Both twins come to term in only 2% of those pregnancies. A singleton is born around 12%



Human embryos can split and become twins early or late in development.

of the time, and in the vast majority of cases both embryos are lost, often without the pregnancy ever being noticed^{2,3}. Some theories suggest that the twin conceptions that survive do so because of a healthy system that is able to support two embryos through implantation and pregnancy.

Founder effect

In Jordan, Reversade used genetic tests to search for shared patterns of single-letter variations in the genomes of the twins that might point to a twinning gene. He found one candidate region, on chromosome four. One of the genes in the region has some promising characteristics: it is conserved through vertebrate evolution, it codes for a protein that is expressed in the blastocyst stage of mice embryos, and its activity drops once the cells that make it have differentiated. Reversade thinks that mutations in the gene or other genes working in the same signalling pathway might have been present in the founders of each twin town and then spread through the population. He says he won't publish the work "until I have the full story, namely the gene, and a possible mechanism".

That story took on new twists during Reversade's visit to India. "The plot thickens," Reversade wrote in an e-mail after returning from Mohammad Pur Umri. "The last pair of monozygotic twins was born 3 weeks ago. There have been close to 55 pairs over the past 30 years. Not all have survived. One mother has had two pairs of monozygotic twins. It is absolutely astounding." Villagers there told him of three sets of twins produced by buffalo that share a pond the village uses for its water supply. Reversade took samples of the water to check

for any substances that could affect reproduction. The beginning of twinning in the village coincides with the establishment of a nearby air force base some 40 years ago, but he could not find any evidence of obvious environmental pollutants. Still, some villagers are convinced there is something in the soil causing the twinning.

At least one related gene hunt has been unsuccessful. In 2004, scientists from the Centre for Cellular and Molecular Biology in Hyderabad collected blood samples from the Indian village to look for genes associated with twinning and appointed a sociologist to survey food intake and social habits. "We have not been able to conclude anything significant," says Lalji Singh, the centre's director. The group did not bother publishing the results.

Reversade hopes that newer technologies — he plans to use next-generation genetic

P. ADHIKARY/EPA/CORBIS

B. MERRIMAN

M. JOHNSON/WELLCOME IMAGES

sequencing machines on the candidate regions in all his subjects — will give him the necessary sensitivity to find genes, but he has a long way to go to convince his peers. Edison Liu, director of the Genome Institute of Singapore, says that inbreeding, which is suspected in each of these villages (although denied by residents of Umri) supports the argument for a genetic contribution to twinning. But, he says, “the observation should be viewed with some scepticism as geographical clustering of rare events often is a statistical fluke”.

Reversade's next stop, probably this summer, will be the village of Linha São Pedro in Brazil, a town predominantly of blond-haired, blue-eyed people founded by German immigrants. In the 1990s, 10% of the births there were twins, and almost half of those were monozygotic. In a book published last year, *Mengele: the Angel of Death in South America*, Argentine journalist Jorge Camarasa argues that Nazi doctor Joseph Mengele was responsible, using techniques he devised to rebuild a master race. The Mengele argument hasn't sold well among scientists and Reversade hopes to prove it wrong once and for all. “That explanation makes it a perfect plot to solve,” he says.

Humans are one of the most proficient mammalian species at multiple monozygotic

births. Armadillos are better. The nine-banded armadillo (*Dasypus novemcinctus*) produces identical quadruplets with every litter from embryos that have split, and then split again. Importantly for Reversade, they also show that multiple births can have a genetic basis

in mammals. Evolution seems to have favoured the genes that contribute to the armadillo's reproductive strategy. Creating multiple embryos from each egg may get around physical constraints imposed by the shape of the armadillo uterus.

It is not clear that reproduction in humans and armadillos would employ the same genes or have arisen for the same evolutionary reasons. Most scientists think of twinning in humans not as an evolutionary

advantage, but as a breakdown of normal function in a female body evolved to carry only one fetus at a time. Some, taking into consideration the 2–3 fold increased risks of congenital anomalies in babies born as monozygotic twins⁴, describe it in terms of risk factors, as they would a disease. “Most people accept that twinning is a failure, not a desired outcome,” says Dianna Payne, an embryologist and visiting researcher at the Mio Fertility Clinic in Yonago, Japan.

Why the system goes awry is not clear.

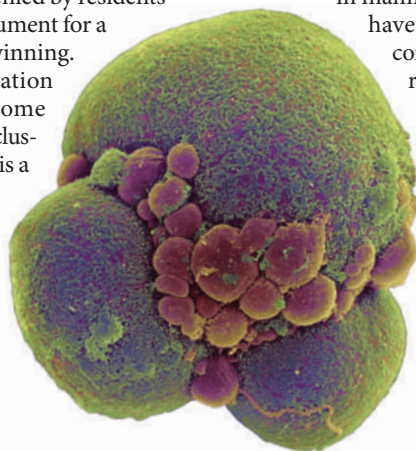
One theory of twinning holds that subtle differences force cells in early embryos to repel one another and establish two separate cell masses. Reversade, hesitantly, speculates that his mutated candidate gene might deprive cells in the embryo of their ability to adhere tightly to each other, resulting in a split.

Hall favours a theory based on the timing of fertilization to explain why humans twin so often. Most mammals recognize the oestrus cycle so that they mate at “the right time”, she says, when eggs have been freshly ovulated. “Humans do it any old time and can easily miss the right time.” She suggests that the outer shell, called the zona pellucida, of an old egg might be more likely to fracture and split the blastocyst in two when it later hatches out⁴. In this scenario, twinning could just be a side-effect of an otherwise successful human reproductive strategy.

Normal genetics

Isaac Blickstein of the Hadassah-Hebrew University School of Medicine in Jerusalem and Louis Keith of Northwestern University in Chicago, Illinois, think that there could be a genetic component to monozygotic twinning, although not one considered mutant or abnormal. They argue that human twinning is the result of “a subpopulation of primary oocytes with an inborn and, as yet, unspecified propensity” for splitting⁵.

There is one thing that all scientists agree on: assisted reproduction causes higher rates of monozygotic twinning — 2 to 12 times higher, according to a review published last year⁶. Ovarian stimulation with drugs, and embryos



Cells in the early embryo may lose their ability to stick together.

MAKING TWINS

The clonal propagation of human twins might sound like the setting for a science fiction novel, but it is already being tested for some types of fertility treatment.

Karl Illmensee, laboratory director at the Genesis Fertility Center in Patras, Greece, thawed and then manually cleaved human embryos at different stages of development then cultured the resulting sections *in vitro*. He found that splitting high-quality embryos at the 6–8 cell stage (pictured) is the most promising way to create viable twin embryos. He split 12 embryos that had been fertilized *in vitro* and developed to this stage, and found that 16 of the resulting 24 embryos developed

into healthy, mature (morula or blastocyst) embryos¹⁰.

These embryos were at the stage at which they could be transferred to the uterus. But Illmensee experimented with ‘triploid’ embryos, which have three, instead of two, sets of chromosomes. These embryos don't usually survive to term and are discarded during *in vitro* fertilization (IVF) procedures.

Illmensee theorizes that the technique would give couples who are low on embryos after IVF a greater chance of becoming pregnant. But would their offspring be healthy? Various animals, including mice, sheep, cattle, goats, pigs and horses have produced



twins after embryo splitting. But efforts by scientists to make genetically identical primates by this technique, for use as a disease model, have resulted in multiple pregnancies but only one live birth, and no twins^{11,12}. Illmensee points to only one attempt to create human

twins in a laboratory, in which the split embryos stopped developing after just a few cell divisions¹³.

Illmensee left the scientific mainstream after some of his experiments were called into question in the early 1980s, including some in which he claimed to clone mice. He says he is being careful to avoid potential ethical objections. He quotes the American Society for Reproductive Medicine's ethics committee report on twinning: “splitting one embryo into two or more embryos could serve the needs of infertile couples in several ways. As long as a couple is fully informed of the risk of such an outcome, there would appear to be no major ethical objection.” **D.C.**

REF.10



Nine-banded armadillos consistently give birth to genetically identical quadruplets.

B. LAVIES/NATIONAL GEOGRAPHIC/GETTY IMAGES

spending an extended period in culture before being transferred are most consistently linked with twinning events. The culture conditions could act directly on the embryo, encouraging it to split, or could harden the zona pellucida so that the embryo is snipped in two as it hatches.

Perhaps the closest researchers have come to watching human twinning is in time-lapse images collected every two minutes during the development of embryos that had been created by *in vitro* fertilization, frozen and then thawed after they were designated as surplus. Payne's team at the Mio Fertility Clinic was surprised to find that 25 of 26 blastocoels, fluid-filled cavities that support the compacted clump of cells that will become the fetus, collapsed at least once. The more frequent and more dramatic the collapse, the less likely the embryos were to survive. Most strikingly, the clump of cells inside two of the collapsed blastocoels split and the two fragments developed as if they were on the path to becoming two separate embryos in different parts of the egg. "The cells are quite sticky at that point and when the blastocoel reinflates itself, some get stuck on the other side," says Payne. She thinks that the culture system, which does not perfectly mimic the uterine environment, might bring about the collapse by causing some cells to die or by weak-

ening cell junctions in the egg's membrane. Such mechanisms could also be taking place in naturally conceived embryos if they were triggered by faulty genes, says Payne. "But we'll never get a camera in there to see it."

Special cameras aren't needed to document another aspect of identical twinning: that the twins differ in appearance, personality and in their propensity to develop disease. "After 50 years of epidemiological work, we cannot answer why there is such divergence in multiple sclerosis, schizophrenia, or type I diabetes between twins," says Arturas Petronis, who has been studying twins at the Univer-

sity of Toronto in Ontario. Last year, Jan Dumanski at the University of Alabama in Birmingham offered a surprising possible answer: that identical twins are not so genetically identical after all. Her group compared 19 pairs of monozygotic twins and found that the individuals within a pair have segments of DNA that are duplicated or deleted⁷. These regions might help to identify the causes of some of the disease discrepancies within the pair.

Other explanations for twin differences might be found outside the genetic sequence. Manel Esteller at the Bellvitge Institute for Biomedical Research in Barcelona, Spain, looked for differences in patterns of histone acetyla-

tion and methylation — 'epigenetic' marks that commonly control gene activity — in monozygotic twins ranging in age from 3 to 74. The younger twins had similar epigenetic marks — but those patterns diverged with age, or as they were exposed to different environmental factors⁸. "When one twin starts smoking, taking drugs or moves somewhere with more air pollution — even for only a year — their epigenetic profile can diverge sharply," says Esteller. "It is very dynamic." Petronis, however, thinks that spontaneous, random epigenetic changes are likely to contribute more than those triggered by the environment.

The epigenetic differences might even start accumulating from day one of development. Researchers at the University of Cambridge, UK, found that the first four cells in a mouse embryo sometimes exhibit differences in histone methylation⁹. Petronis suggests that diverging epigenetic profiles might be what drives the cells to split into twins in the first place. His work on monozygotic twins has shown that the genomic loci that differ the most epigenetically are those involved in cell-division processes, which "may reflect an early developmental discordance as one of the hypothetical reasons of twin formation".

If that is the case, Reversade might not ever find a twinning gene. But that is not going to stop him from packing his bag for the twin town in Brazil. The armadillo gives him confidence that there are twinning genes to be found, as does a strain of zebrafish in which a mutation can generate a double head like that of conjoined twins. And so does the Jordanian family which, he says, "cannot be explained by these epigenetic phenomena but can be by conventional genetics".

What will he do if his results don't show a gene? "Publish them," he says simply. ■

David Cyranoski is *Nature's* Asia-Pacific correspondent.

- Lewis, C. M., Healey, S. C. & Martin, N. G. *Am. J. Med. Genet.* **61**, 237–246 (1996).
- Boklage, C. E. *Int. J. Fertil.* **35**, 75–94 (1990).
- Boklage, C. E. The frequency and survival probability of natural twin conceptions. In *Multiple Pregnancy: Epidemiology, Gestation, and Perinatal Outcome* (1995).
- Hall, J. G. *Lancet* **362**, 735–743 (2003).
- Blickstein, I. & Keith, L. *Twin Res. Hum. Genet.* **10**, 394–399 (2007).
- Aston, K. I., Peterson, C. M. & Carrell, D. T. *Reproduction* **136**, 377–386 (2008).
- Bruder, C. E. *et al. Am. J. Hum. Genet.* **82**, 763–771 (2008).
- Fraga, M. F. *et al. Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).
- Torres-Padilla, M.-E., Parfitt, D.-E., Kouzarides, T. & Zernicka-Goetz, M. *Nature* **445**, 214–218 (2007).
- Illmensee, K., Levanduski, M., Vidall, A., Husami, N. & Goudas, V. T. *Fert. Steril.* doi:10.1016/j.fertnstert.2008.12.098 (in the press).
- Chan, A. W. S. *et al. Science* **287**, 317–319 (2000).
- Mitalipov, S. M., Yeoman, R. R., Kuo, H. C. & Wolf, D. P. *Biol. Reprod.* **66**, 1449–1455 (2002).
- Hall, J. L. *et al. Fertil. Steril.* **61**, 51 (1993).

"Most people accept that twinning is a failure, not a desired outcome."
— Dianna Payne

CORRESPONDENCE

Polar science: global partnership to work on data sharing

SIR — Your Editorial ‘The way ahead for polar science’ (*Nature* **457**, 1057; 2009) correctly points out the critical need for a sustained archive of the data collected during International Polar Year (IPY). As you mention in your News Feature ‘In from the cold’ (*Nature* **457**, 1072–1077; 2009), “the IPY data-management committee is sorting out the options for setting up a fully integrated data-sharing system”. The primary mechanism for this data sharing is the International Polar Year Data and Information Service (IPYDIS).

The IPYDIS is a global partnership of data centres, archives and networks, which together ensure proper stewardship of IPY and related data (<http://ipydis.org>). We seek to create a union catalogue that provides access to all IPY data distributed in archives around the world.

As you emphasize, sustained support for this activity is a critical challenge for nations and for the international science community. Australia, Canada, China, the Netherlands, Norway, Russia, Sweden, the United Kingdom, the United States and others have already started setting up a system, but the work to present the IPY data legacy has only just begun.

Although the IPY data policy encourages more comprehensive and more timely release of data, data sharing is still a challenge for polar science. We need to build on our success and extend the effort to a broader community. To this end, the International Council for Science has established a new project, ‘The Polar Information Commons: establishing the framework for long-term stewardship of polar data and information’. The council’s committee on data for science and technology, CODATA, leads this initiative to establish a

sustainable long-term framework for the preservation of and access to polar data.

We encourage a ‘commons’ approach, as used in other disciplines, that recognizes information as a common societal benefit. Historically, society has managed both to share and administer common property, such as fisheries and telecommunications bandwidth, but to sustain these resources requires ongoing societal attention. This approach can strengthen incentives for scientists, research institutions and nations to contribute and document data, reduce barriers to data sharing and provide a focal point for data assessment by the community.

Mark A. Parsons National Snow and Ice Data Center/World Data Center for Glaciology, University of Colorado, 449 UCB, Boulder, Colorado 80309-0449, USA
e-mail: parsonsm@nsidc.org

Polar science: bid for freely accessible biodiversity archive

SIR — The Southern Ocean is of unique ecological, biogeographic and political interest, as noted by your Editorial ‘The way ahead for polar science’ (*Nature* **457**, 1057; 2009). Given the rate and potential catastrophic impact of global change on the Antarctic marine ecosystem, we agree that it is vital to have a mechanism for the exchange of relevant scientific information. Information about Antarctic marine biodiversity has to be widely published, instantly accessible and thoroughly checked, in order to enable timely, science-based management.

An example of the kind of initiative you call for is the Marine Biodiversity Information Network of the Scientific Committee on Antarctic Research (SCAR-MarBIN). Since 2005, the network’s web portal has compiled and managed information on Antarctic

marine biodiversity (www.scarmarbin.be). It is home to the first complete Register of Antarctic Marine Species, a fully operable, browsable and searchable online species list, maintained by more than 70 expert taxonomic editors. The register offers access to information on more than 15,000 taxa and contributes to the World Register of Marine Species and to the Catalogue of Life.

Through a mapping system, SCAR-MarBIN also allows baseline data on marine organisms to be visualized and downloaded. This offers, for the first time, a way to quantify the diversity and distribution of Antarctic marine life and to measure how, when and where it has been studied. This information is also made available through larger initiatives, such as the Ocean Biogeographic Information System or the Global Biodiversity Information Facility.

The data published in SCAR-MarBIN constitute a benchmark for detecting responses to global change, enabling the scientific community to highlight key areas that require investigation or protection. Currently focusing on biogeographic data, SCAR-MarBIN is developing tools to improve its usefulness for conservation and sustainable management.

As part of the legacy of International Polar Year, various initiatives are collaborating to document biodiversity in the Southern Ocean, the ecosystem services it delivers and its evolution in the light of global change. Creating a freely accessible archive should be considered a top priority, with funding from all the International Polar Year nations relative to their economic strength. It is a matter of responsibility towards future generations.

If they are successful, initiatives including SCAR-MarBIN will safeguard the legacy of these efforts, helping to preserve Antarctica as was designated by the Antarctic Treaty: as “a natural

reserve, devoted to peace and science”.

Bruno Danis Royal Belgian Institute of Natural Sciences, 29 rue Vautier, 1000 Brussels, Belgium
e-mail: bruno.danis@scarmarbin.be
Huw Griffiths British Antarctic Survey, Natural Environment Research Council, Cambridge CB3 0ET, UK

Canadian government reaffirms support for science and discovery

SIR — You report researchers’ concerns about the Canadian government’s support for science in two recent News stories (*Nature* **457**, 646; 2009 and *Nature* **458**, 393; 2009). As Minister of State for Science and Technology, I can say that, despite the global economic situation, the government of Canada remains committed to innovation and discovery. We have increased funding to researchers, both in universities and in the private sector.

In the past three years, for example, we have significantly increased the budgets of federal granting councils, increased scholarships through the Canada Graduate Scholarships Program, and increased the Industrial Research Assistance Program for small and medium-sized businesses. The Budget 2009 announcements include Can\$750 million (US\$590 million) for the Canada Foundation for Innovation to attract and retain world-leading researchers, and a Can\$2-billion infrastructure programme. The government has also put in place two five-year funding agreements with Genome Canada that are worth Can\$240 million, to support large-scale, world-class research.

Your readers should therefore rest assured that the government of Canada will continue to fund research for the benefit of all scientists and Canadians.

Gary Goodyear Industry Canada, 235 Queen Street, Ottawa, Ontario K1A 0H5, Canada
e-mail: gary.goodyear@ic.gc.ca

DAVID PARKINS

Dormant microbes: scouting ahead or plodding along?

SIR — In his Essay 'Microbial awakenings' (*Nature* **457**, 1083; 2009), Slava Epstein proposes that microbial cells stochastically revert from a state of dormancy to a growing state, allowing a clonal population of dormant cells to exploit rare and transient conditions in an environment with unpredictable periods of feast and famine. This 'scout' strategy is compatible with some — but not all — explanations for the general unculturability of environmental microbes.

Other explanations include oxidative stress and substrate-accelerated death on transfer to the laboratory environment, and slow-growth strategies. If most microbes in energy-limited and spatially heterogeneous environments are slow-growing species ('K-strategists'), they may be permanently prepared for a very slow but steady existence under nutrient limitation. Rapid growth won't occur because the cells would not then be ready for the end of a brief nutrient flush. Microbes using the K strategy are always prepared for resource exhaustion. As a result, they grow only slowly in laboratory culture, as noted for many 'unculturable' soil bacteria.

Other microbes, rarer in nutrient-poor environments, respond rapidly to nutrient flushes and grow at much higher rates when resources are plentiful. However, these microbes ('r-strategists') must re-enter a dormant state before such a nutrient flush is exhausted. This transition is an active and energy-requiring process, and cell death is rapid if the transition cannot be completed. Once such cells are in a dormant state, they wait for the next nutrient flush while slowly losing viability.

Therefore, Epstein's scout theory could apply to



opportunistic *r*-strategists, but not so easily to *K*-strategists.

Peter H. Janssen AgResearch Limited, Grasslands Research Centre, Palmerston North 4442, New Zealand
e-mail: peter.janssen@agresearch.co.nz

Dormant microbes: time to revive some old ideas

SIR — A number of researchers have already investigated, and demonstrated in detail, some of the propositions put forward in Slava Epstein's Essay 'Microbial awakenings' (*Nature* **457**, 1083; 2009).

For instance, Epstein says "researchers could attempt to identify signalling compounds used by scouts, if they exist, and use these molecules to induce growth of dormant kin". This has been investigated by G. V. Mukamolova and colleagues (*Proc. Natl Acad. Sci. USA* **95**, 8916–8921; 1998).

The proposition that "single cells could be cultivated, allowing them enough time for scout production unimpeded by faster growing species" has also been demonstrated (A. S. Kaprelyants *et al.* *FEMS Microbiol. Lett.* **115**, 347–352; 1994). The phenomena of transient dormancy and culturability — which Epstein says "contrasts with scientific expectations" — have been studied and reviewed extensively (see, for example, G. V. Mukamolova *et al.* *Adv. Microb.*

Physiol. **47**, 65–129; 2003).

Evidently ideas, as well as microbes, may remain dormant for some time.

Douglas Kell School of Chemistry and Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M1 7DN, UK
e-mail: dbk@manchester.ac.uk

The reading list for the Essay 'Microbial awakenings' (Nature 457, 1083; 2009) was published online, but Nature inadvertently omitted to refer to this list in the PDF version.

Technology takes on deadlines for fetal human rights

SIR — The view expressed in Patricia Pranke and João Carlos Silveiro's Correspondence 'Human rights cannot cover cells that were never in the womb' (*Nature* **458**, 147; 2009) is one of the standard components in a belief system that detaches personhood from its genetic basis. (Others include birth, viability, ability to feel pain, and various developmental milestones.)

The issue is clouded by our prejudices about what looks human and by the way fetal development seems to be a continuum. This belief system is weakened by its arbitrariness, which is being exposed as technology pushes back the boundaries. The implantation argument will become irrelevant when 'artificial wombs' make ectogenesis possible. There is no reason to expect this will not be

"Evolution stopped when the first caveman spoke."

Tom Wolfe, page 837

developed in the future, whether governments try to prohibit it or not — as has been predicted for many years (Canadian Royal Commission on New Reproductive Technologies *Proceed With Care*; Canadian Gov. Publ., 1993).

Michael C. Wendt Department of Genetics, Washington University Medical School, Box 8501, St Louis, Missouri 63108, USA
e-mail: mwendt@wustl.edu

More acclaim for Darwin's theory of sexual selection

SIR — In this bicentennial year of Charles Darwin's birth, it is gratifying to see *Nature* devoting a wealth of comment to the great man and to evolution. But like much other scholarly perspective, your emphasis is on natural selection, the first of Darwin's great ideas.

Yet his second great thesis, on sexual selection, is in some ways his greatest work, because it is far less obvious than natural selection. After all, Alfred Russel Wallace also discovered natural selection, but no one else envisaged anything like sexual selection, which, even after Darwin's exposé, remained something of an anathema. Sexual selection is responsible for some of the most spectacular behaviours and characters in nature (picture the peacock's tail), but it was largely ignored during the modern synthesis and only really became mainstream in the 1970s.

Since then, we have come to realize the importance of sexual selection for much of biology, from driving rapid molecular evolution to speciation, the subject of Darwin's best-known work, *On the Origin of Species*. In this celebratory year, we need also to acknowledge the importance of this historically maligned mechanism of evolution.

David J. Hosken School of Biosciences, University of Exeter, Cornwall Campus, Penryn, Cornwall TR10 9EZ, UK
d.j.hosken@exeter.ac.uk

ESSAY

Can evolution explain how minds work?

Biologists have tended to assume that closely related species will have similar cognitive abilities.

Johan J. Bolhuis and **Clive D. L. Wynne** put this evolutionarily inspired idea through its paces.

Darwin's theory of evolution by natural selection is broadly accepted among biologists, but its implications for the study of cognition are far from clear. Few within the scientific pale would argue against the proposition that life on Earth has evolved and that this general principle can be extended to the process of thought. But in taking an evolutionary approach, biologists have tended to assume that species with shared ancestry will have similar cognitive abilities, and that the evolutionary history of traits can be used to reveal how we and other animals perform certain mental tasks. A closer analysis suggests things aren't so simple.

In *The Descent of Man*, Darwin proposed that there is "no fundamental difference between man and the higher mammals in their mental faculties" on the basis of his belief that all living species were descended from a common ancestor. He also suggested that "there is a much wider interval in mental power between one of the lowest fishes ... and one of the higher apes, than between an ape and man". To support his argument, he outlined cases in which forerunners of human intelligence could be found in "higher mammals", including "similar passions, affections, and emotions, ... [such as] ... jealousy, suspicion, emulation, gratitude, and magnanimity".

Darwin's reports of "a sense of humour ... wonder and curiosity" or "the association of ideas, and reason" in animals may seem far-fetched, but many contemporary researchers do not shy away from using similar anthropomorphic language in their interpretation of animal behaviour. Over the past two decades, researchers have reported that chimpanzees can empathize with other members of their species, and that they reconcile and even console each other after conflicts. Monkeys and apes have been credited with a sense of fairness and aversion to inequity and, in the case of apes, an awareness of the mental states of others — in other words, a theory of mind.

A closer look at many of these studies reveals, however, that appropriate control conditions have often been lacking, and simpler explanations overlooked in a flurry of anthropomorphic overinterpretation. For instance, capuchin monkeys were thought to have a sense of fairness because they reject a slice of cucumber if they see another monkey in an adjacent cage, performing the same task, rewarded with a more-sought-after grape. Researchers

interpreted a monkey's refusal to eat the cucumber as evidence of 'inequity aversion' prompted by seeing another monkey being more generously rewarded. Yet, closer analysis¹ has revealed that a monkey will still refuse cucumber when a grape is placed in a nearby empty cage. This suggests that the monkeys simply reject lesser rewards when better ones are available.

Such findings have cast doubt on the straightforward application of Darwinism to cognition. Some have even called Darwin's idea of continuity of mind a mistake².

One solution fits many

Laboratory studies of a number of species performing a wide range of tasks indicate that different species may have arrived at similar solutions to cognitive problems because they have experienced similar selection pressures, not because they are closely related. In other words, evolutionary convergence may be more important than common descent in accounting for similar cognitive outcomes in different animal groups.

For example, we now know that birds are capable of feats that match or even exceed those reported in monkeys and apes. Rooks, for example, rub their bills together after one of them has been involved in a confrontation with another bird. Equivalent stroking and embracing in chimpanzees would be labelled 'consolation'. The self-directed pecking that magpies show when they are put in front of a mirror after a mark has been placed on their body is similar to the reactions seen in apes given the same treatment. In magpies, this behaviour has been interpreted as evidence for some degree of self-recognition. But in apes, the same behaviour has been thought to indicate a deeper level of self-consciousness. Caledonian crows outperform monkeys in their ability to retrieve food from a trap tube — from which food can be accessed only at one end. The crows can also work out how to use one tool to obtain a second with which they can retrieve food, a skill that monkeys and apes struggle to master.

Researchers have tried for decades to teach apes some form of language, be it by using visual symbols or gestures. But linguists generally agree that the resulting efforts made by chimps and bonobos don't qualify as language³. One of the prerequisites for language is being able

to imitate sounds that are created by someone else. Our primate cousins show no inclination to do this. Yet many parrots and songbirds are striking vocal mimics. Furthermore, the way that they learn to sing is not unlike how human infants learn to speak. Both children and the chicks of parrots and songbirds learn many of their vocalizations during a sensitive period early in life. They also undergo a transitional period during which their attempts to speak or sing increasingly come to resemble those of adults. Recent studies even suggest that starlings can identify certain syntactic features of sound patterns that non-human primates miss.

The appearance of similar abilities in distantly related species, but not necessarily in closely related ones, illustrates that cognitive traits cannot be neatly arranged on an evolutionary scale of relatedness.

The wrong question

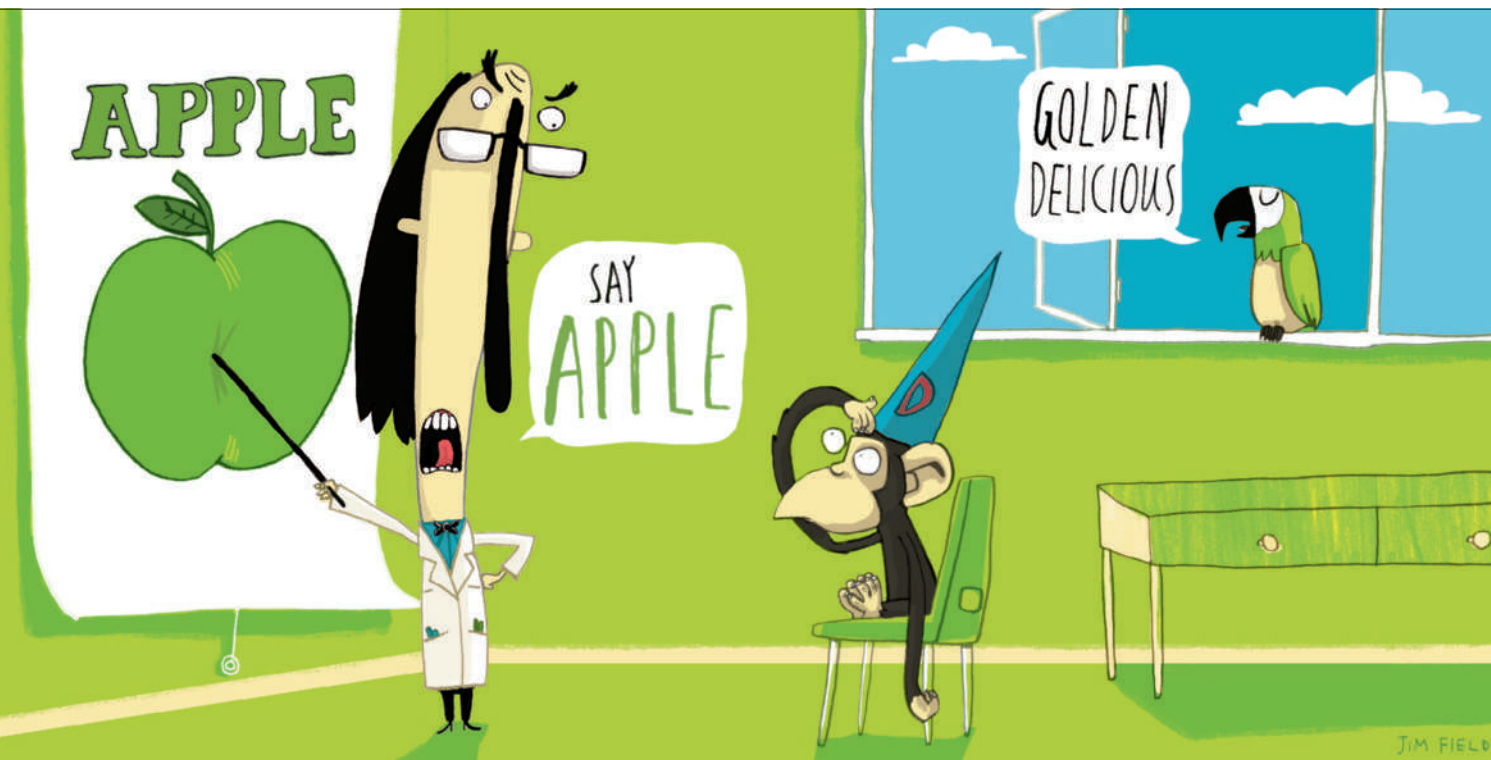
The difficulty of not knowing whether shared ancestry or convergence accounts for similar cognitive outcomes in different species is not the only problem with applying an evolutionary approach to cognition. Another major stumbling block is that it is extremely difficult, if not impossible, to identify the factors that originally drove the emergence of contemporary animal and human traits.

According to Leda Cosmides and John Tooby⁴, protagonists of evolutionary psychology, "Our modern skulls house a Stone Age mind". Indeed, this assumption — that the human mind evolved as a result of selection pressures faced by our Stone Age ancestors — underpins the field. Thus, the tendency of

modern humans to spontaneously fear spiders rather than cars, which are far more dangerous, is thought to stem from the prevalence of poisonous arachnids, rather than dangerous driving, during the Pleistocene.

This approach overlooks the importance of culture in shaping the human mind. It also assumes that all traits evolve as a result of natural selection, whereas they may be inconsequential, or by-products of selection acting on some other trait. However, the most serious problem with this perspective is that cognitive traits of past generations leave little trace in the fossil record. Without being able to reconstruct the mind of our hunter-gatherer predecessors, we can only

"Some have even called Darwin's idea of continuity of mind a mistake."



J. FIELD guess at the selection pressures they faced⁵.

Last but not least, evolutionary analyses have been used to tackle questions for which they are ill suited. Nobel laureate Niko Tinbergen famously considered evolution to be one of the four great problems in behavioural biology. He pointed out that questions concerning the development, function and mechanistic causes of behaviour deserve equal weight⁶. Many of the questions about cognition asked by psychologists and behavioural biologists concern the underlying causal mechanisms. How does the dog open the gate? How does the parrot imitate human vocal sounds? Evolutionary analyses, however, because they are analyses of history, cannot uncover how an animal achieves a particular feat.

Take, for example, food storing in closely related bird species. The marsh tit stores seeds in tree bark or in the ground, and is able to retrieve them several days later. Its close relative the great tit, on the other hand, doesn't store at all. Such differences between species in their reliance on food stores have led researchers to suggest that the ability to remember the location of buried food involves an adaptively specialized spatial memory and brain structure, the hippocampus. To verify this idea, researchers have attempted to show that spatial memory and the size of the hippocampus vary between species depending on the degree to which they store food.

This is an interesting evolutionary hypothesis. But the suggestion that a specialized memory and hippocampus are needed specifi-

cally for storing food is not supported by the evidence⁶. Comparative studies of storing and non-storing bird species have failed to reveal a consistent relationship between the size of the hippocampus and food-storing capability. Moreover, food-storing species in general do not perform any better in spatial-memory laboratory tasks than non-storers⁶.

Had it been discovered, a relationship between hippocampus volume and food storing would have hinted that a larger hippocampus underpins a superior spatial memory. Even then, a causal analysis would be needed to test this hypothesis. It could be, for example, that the hippocampus affects another trait that differs between storers and non-storers. Questions about the causal underpinnings of behavioural differences can be elucidated only with a causal analysis, not through reconstructing evolutionary history.

Theory and practice

Clearly, functional and evolutionary questions are intertwined, as are questions of causation and development. It is unclear, however, what an analysis of the evolutionary history of cognitive behaviours could add to our understanding of how they work, even if such an analysis were possible. At most, an evolutionary interpretation could provide clues to the underlying mechanisms responsible — but such clues would have to be verified using controlled experiments.

We are not suggesting an abandonment of Darwin's insights. Rather, we call for care in

their application. When reconstructing the evolutionary history of cognitive traits, there is no a priori reason to assume that convergence will be more important than common descent or vice versa. In addition, evolutionary theory may suggest hypotheses about the mechanisms of cognition, but it cannot be used to actually study these mechanisms.

As long as researchers focus on identifying human-like behaviour in other animals, the job of classifying the cognition of different species will be forever tied up in thickets of arbitrary nomenclature that will not advance our understanding of the mechanisms of cognition. For comparative psychology to progress, we must study animal and human minds empirically, without naive evolutionary presuppositions. ■

Johan J. Bolhuis is in the Department of Biology and Helmholtz Institute, Utrecht University, the Netherlands. **Clive D. L. Wynne** is in the Department of Psychology, University of Florida, Gainesville, USA.
e-mails: j.j.bolhuis@uu.nl; wynne@ufl.edu

1. Wynne, C. D. L. *Nature* **428**, 140 (2004).
2. Penn, D. C., Holyoak, K. J. & Povinelli, D. J. *Behav. Brain Sci.* **31**, 109–178 (2008).
3. Pinker, S. *The Language Instinct: How the Mind Creates Language* (W. Morrow, 1994).
4. Cosmides, L. & Tooby, J. *Evolutionary Psychology: A Primer* (1997); available at www.psych.ucsb.edu/research/cep/primer.html.
5. Richardson, R. C. *Evolutionary Psychology as Maladapted Psychology* (MIT Press, 2007).
6. Bolhuis, J. J. & Verhulst, S. *Tinbergen's Legacy: Function and Mechanism in Behavioral Biology* (Cambridge Univ. Press, 2009).

BOOKS & ARTS

A measure of importance

Astronomer François Arago of the Paris Observatory defied war, disease and death to survey the meridian running through his city — and helped define the metric system we use today, explains **Andrew Robinson**.

Full Meridian of Glory: Perilous Adventures in the Competition to Measure the Earth

by Paul Murdin

Springer: 2009. 187 pp. £15.99

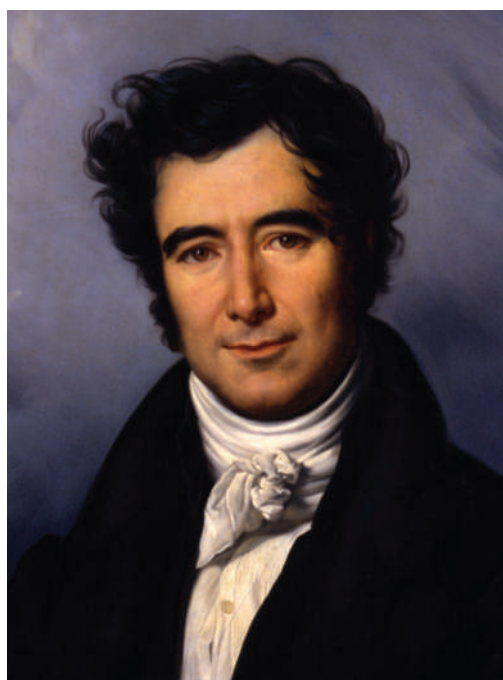
Astronomer Paul Murdin began his career in 1963 at the United Kingdom's Royal Greenwich Observatory. Proud of his employer's history, going back to the observatory's foundation by King Charles II in 1675, Murdin became curious about the prime meridian running through Greenwich that, in 1884, came to define Greenwich Mean Time. He learned, however, that its roots lay not in Britain but in seventeenth- and eighteenth-century France, intertwined with the French Revolution. The accurate scientific measurement of the Paris meridian also formed the basis of the nineteenth-century metric system and its modern form, the *Système International* (SI). Murdin's book *Full Meridian of Glory*, its title taken from Shakespeare's *Henry VIII*, is the outcome of his long fascination with the Paris Observatory, founded by Louis XIV in the late 1660s. Murdin's interest is reinforced by his regular professional visits to the Paris headquarters of the European Space Agency.

Others have trodden this path before, notably Ken Alder in his prizewinning *The Measure of All Things* (Little, Brown; 2002), a historian of science mentioned only once by Murdin. Although a better narrative writer than Murdin, Alder focuses on the 1790s, whereas Murdin ably covers four centuries. Alder also lacks the enthusiasm and knowledge of a working scientist.

Murdin knows what it is like to make night-time observations in sub-zero conditions. He would empathize with the French savants who, in 1734–44, laboriously surveyed and compared the length of a degree of latitude in polar Lapland with that in the equatorial Andes, to prove whether Earth was flattened or elongated at the poles. "I can speak of the pain from personal experience, having left a ring of skin from around my right eye frozen to an eyepiece when I withdrew my head from a telescope in an upstate New York observatory," Murdin writes. He recounts how French astronomer Pierre Charles Le Monnier "had a similar experience in Lapland when his tongue froze to a silver cup from which he was

drinking brandy". And when Murdin includes a full-page photograph of NASA's notorious Mars Climate Orbiter — lost in space in 1999 as a result of a muddle between the imperial units used in its design and the metric units used in its operation — the past and present are seamlessly integrated.

The same cannot be said for the main text. Murdin draws on an extensive and distinguished cast of scientists, including the four key figures of the Cassini dynasty and their nemesis Isaac Newton, whose theory on gravity



François Arago was influential in both science and politics.

was opposed by the elder Cassinis. The Cassini family reigned at the Paris Observatory for more than a century, and mapped France through their combined efforts. Also mentioned are the scientists who surveyed the Paris meridian at the turn of the nineteenth century — in particular, Jean-Baptiste-Joseph Delambre, Pierre Méchain and François Arago. Instead of weaving them into the main text, Murdin mostly discusses their lives in numerous thumbnail sketches set apart in paragraphs of small type. This typographical separation of the biographical from the scientific makes for an indigestible and sometimes confusing read, even though the sketches are informative and often enjoyable.

Fortunately, Arago, clearly a hero to Murdin, gets a lively chapter of his own. Arago became director of the Paris Observatory in 1843. In 1848, he was briefly the minister of war and of the navy, and even the head of state for 46 days during that Revolutionary year. He is known for applying the wave theory of light — proposed by his friend Thomas Young — to stellar aberration, and for suggesting the crucial test between wave and particle theory by comparing the speed of light in water and air. But Arago's journey to survey the mountains of Spain during the war with France in 1806–09, following the death of Méchain from malaria, was both significant science and colourful adventure. Running the gauntlet of isolation, bad weather, equipment failure, brigands, homicidal Spanish mobs and a British blockade of French ports, Arago completed his measurement of the meridian at Majorca in the Balearic Islands. At one point, he had the strange experience of reading in a newspaper about his own execution by hanging, in which he had apparently met death heroically.

In an unusual final chapter, Arago returns as Murdin gives a guided walk through Paris following the trail of more than 100 bronze discs set into the pavement, which carry merely the inscription 'ARAGO' and the compass direction letters N and S. They mark the line of the Paris meridian. The discs were created by a Dutch conceptual artist, Jan Dibbets, winner of an open competition to honour Arago on the bicentenary of his birth in 1986. "The idea was that the people of Paris would accidentally come across the medallions, wonder about them and thus discover Arago," Murdin explains.

At the Louvre, three Arago medallions traverse the Denon Wing, and five others run across the Cour Carrée behind the glass pyramid. In Dan Brown's *The Da Vinci Code*, Murdin informs us, the novel's hero is drawn to the pyramid by the discs. Yet as this quirky book demonstrates, the discs are not mysterious markers but a celebration of French science at its most rational, symbolized by Arago. ■

Andrew Robinson is a visiting fellow at Wolfson College, Cambridge CB3 9BB, UK. He is the author of *The Story of Measurement* and *The Last Man Who Knew Everything*, a biography of Thomas Young.

e-mail: ar471@cam.ac.uk

C. STEUBEN/OBSERVATOIRE DE PARIS

Decisions, decisions...

How We Decide/The Decisive Moment

by Jonah Lehrer

Houghton Mifflin Harcourt/Canongate:

2009. 320 pp./304 pp. \$25/£16.99

"I was flying a Boeing 737 into Tokyo when the left engine caught on fire," begins Jonah Lehrer in his captivating book on decision-making. What should a pilot do? The right decision will make the difference between life and death. Luckily, Lehrer was piloting a simulator and could replay his decisions. But how do you decide when you don't get another chance? *How We Decide* uses the lessons of decision-making research to help us make better choices.

Lehrer asks if good decisions are always reached by extensive rational deliberations, or if we should sometimes rely on our automatic, emotional reactions. When is it best to use our gut instincts and when should we stop and reason? He organizes his answer around vivid anecdotes, each highlighting key concepts from psychology and behavioural economics. Lehrer is an expert storyteller, often using dramatic examples from aviation to highlight how rapid gut reactions averted one crash, whereas emotional control and reasoned, outside-the-box thinking by pilots let them survive in others.

Sometimes it is best to stop thinking altogether. When trying to make optimal choices based on a large number of attributes, such as picking a car or a house, people are less accurate when left to ponder for a few minutes than when they are distracted for a similar period. One interpretation is that conscious thought is at a disadvantage when processing many attributes in parallel, owing to limitations in capacity. Such examples remind us of the limits of our deliberative decision-making processes.

Lehrer considers any decision that is not executed according to the rules of logic to be "irrational". But in many cases these only show the limitations of laboratory studies. The book fails to note relevant perspectives from behavioural economics and game theory, according to which many emotions are rational evolutionary adaptations that have essential roles in cooperative behaviours and social interaction. More problematic is Lehrer's insistence on attributing decisions to either an emotional brain or a rational one. There is no evidence that the brain has distinct and opposing emotional and rational regions. A case in point is the dopamine system. Although described in the book as emotional, our current understanding is that it is rational: dopamine neurons signal a simple and computationally well-defined variable — the reward prediction error — and broadcast it



A fast-thinking pilot saved lives by landing his plane on the Hudson River in New York.

throughout the brain. This simplification of the complex function of the dopamine system is a major success story of neuroscience. As it turns out, the process by which these neurons signal errors may be suboptimal, but that does not make it irrational or emotional.

Lehrer has a tendency to present a caricature view of brain processing. For example, the nucleus accumbens "might want the George Foreman grill, but the insula knows that you can't afford it, or the prefrontal cortex realizes that it's a bad deal. The amygdala might like Hillary Clinton's tough talk on foreign policy, but the ventral striatum is excited by Obama's uplifting rhetoric."

Such neurobabble has unfortunately become commonplace in science journalism. Lehrer is following a well-established lingo — and some blame lies with the scientific community for tolerating, if not encouraging, this trend. Until

recently we could not study a functioning human brain. All we had were the accidental lesions of patients whose tragic stories illuminated the functions of their missing brain regions. Recent advances in neuroscience dramatically changed this situation — but in this book the visual appeal of neuroimaging studies is often used as a substitute for the lessons learned. It is exciting to observe the nucleus accumbens 'lighting up' in a brain scan when a reward is provided, but this is not an explanation for 'wanting' and misses the subtler yet more enlightening findings about reward processing in the brain.

Lehrer aims high by applying lessons from the neuroscience of decision-making to everyday choices. He provides valuable — if obvious — advice, such as relying on gut instinct for unimportant decisions but on reason when faced with new problems. But we were promised insights from neuroscience, not age-old wisdom.

Neuroscientists must grapple with how to communicate subtle conclusions in a world of headline-news reporting. Popular books such as this one could serve this goal of translating science for the public. But by smoothing out the rough edges of our knowledge, Lehrer leads the reader astray. Among his well-told dramatic stories the science often gets lost. Yet he is a gifted writer, so I hope for his next project he will dig deeper, embrace complexity and make it his job to report from the cutting-edge of scientific knowledge. ■

Adam Kepecs is assistant professor of neuroscience at Cold Spring Harbor Laboratory, New York 11724, USA.
e-mail: kepecs@cshl.edu

Songs on the brain

Rock-It Science

Concert at the Highline Ballroom, New York
3 March 2009

Swapping his microscope for a microphone, New York University neuroscientist Joseph LeDoux (pictured, right) took to the stage of the Highline Ballroom, a Manhattan music club. Leading his folk rock band, LeDoux articulated each line of his science-inspired songs with conviction, accompanied by electric guitar. His group, The Amygdaloids, featured heavily at *Rock-It Science*, a concert held last month to raise funds for research into sensory-processing disorders.

The charity concert, the brainchild of LeDoux and co-organized with psychologist Jennifer Jo Brout of the Sensation and Emotion Network in New York, mixed scientists with



professional musicians, and was headlined by singer-songwriter Rufus Wainwright.

LeDoux, who studies the biological mechanisms underlying emotional memory, sees rock 'n' roll as a perfect vehicle for inspiring interest in neuroscience. "All our songs have a mind-brain angle. We believe we can interest someone in tough concepts through music."

B. MCDERMID/REUTERS

S. LOVEKIN/GETTY

The Amygdaloids performed tunes from their new album, *Brainstorm*, some of which explore how love can disrupt normal thought processing. In *Crime of Passion*, a jealous lover's rage momentarily overwhelms his rational thinking: "Sentenced to death for a crime I did commit/I couldn't stop/I did it in a fit/of anger and pain."

Cognitive psychologist, vocalist and guitarist Daniel Levitin of McGill University in Montreal, Canada, sang a scientific homage to the 1989 hit *Wicked Game* in which Chris Isaak croons "No, I don't want to fall in love." Levitin's version,

I Don't Want My Brain Cut In Two, refers to a drastic procedure once used to treat epilepsy.

Levitin says his experiences in the recording industry now inform his work in the lab. "I was always astonished by how long it takes to make a three-minute song, and how many times you have to record to get it right." He explains that he has to be similarly patient when repeating experiments or redrafting a research paper.

Other scientist performers included neuroscientist and classical guitarist David Sulzer, known on stage as Dave Soldier, and

geneticist Pardis Sabeti, of Boston-based band Thousand Days. Among the professional musicians was guitarist Lenny Kaye, who has recorded with punk singer Patti Smith, and Twisted Sister's Dee Snider.

Wainwright held nothing back in his set. He professed he "failed every science course [he] ever took", but supported the charity gig because "scientists have become the new oppressed people, and I'm always there for the oppressed."

Roxanne Khamsi is news editor at *Nature Medicine* in New York.

Flashes of cosmic brilliance

Tim Otto Roth's minimalist art installation reflects the complexity of cosmic radiation, explains **Martin Kemp**.

Daylight is fading over a grass-covered field. Looking like extra-large bee hives, 252 white cabins on short stilts are regularly distributed across its vast expanse of 44,000 square metres. A larger 'house' occupies the space of the central 4 units. At the centre of each square block of 16 cabins is set a two-tier, highly reflective, cylindrical device.

As darkness envelops the array, short-lived flashes of light erupt unpredictably from the 16 devices. Sometimes one launches alone. At other times they evoke straight lines, or neat clusters. Or just flicker here and there. The flashes are so rapid and apparently random that our eyes search in vain for a repeated pattern. Logic appears to be defied. What are we witnessing? A high-tech scientific experiment? Some sci-fi setting for a war of the worlds? An evocation of a night-time bombing raid?

We are, in fact, watching a minimalist-style art installation by German artist Tim Otto Roth. His *Cosmic Revelation* (see <http://tinyurl.com/cosrev>) presents a sensory experience of the cosmic radiation detected by the KASCADE project at the Centre for Elementary Particle and Astroparticle Physics at the Karlsruhe Institute of Technology, Germany. Roth's installation is both an artwork and an act of science communication. And it interacts in an unexpected and compelling way with the systematic recording of cosmic phenomena.

The cabins contain detectors for the elementary particles that move at speeds close to that of light and crash incessantly into our atmosphere. These multiple collisions initiate cascades of millions of new particles, and KASCADE measures those that have energies



Cosmic-ray showers trigger a light display at the KASCADE array.

of around 10^{14} – 10^{17} electronvolts (eV; one eV is equivalent to the energy acquired by an electron falling through a difference in potential of one volt). Together with other detector units distributed across the research centre, the KASCADE-Grande experiment can detect energies as high as 10^{18} eV. The results are recorded in a central laboratory by neat grids of LEDs. KASCADE's scientists can view the topographical display of data received by all 252 cabins live on the Internet.



Roth's 16 strobe-light units are in themselves pieces of high-tech sculpture that come into their own as beautiful objects during daylight hours. A 1,500-watt linear flash tube in each unit is protected from the weather by the upper disc. A convex mirror is set into the lower disc so as to project the pulses of light radially from the cylinder's open sides.

These 'cosmic mirrors' carry the visual outputs of the laboratory apparatus back into the field of the outdoor detectors, transforming the energies of the unseen into wavelengths accessible to human sight. Thus transformed,

the particles' energies are transmitted back into the space through which they have cascaded.

When Roth categorizes his work as minimalist, he is referring to the reductive mode in art and music that arose in the late 1960s. It relies on the stripping down of a composition into basic units, often regularly repeated and frequently mathematical in form. From the repetition and regularity, paradoxical variousness and complexity emerge when the viewer moves or the lighting changes. Reflective surfaces are particularly favoured.

The generation of complexity from simplicity parallels those physical phenomena that fall under the embrace of chaos theory, and mirrors the reverse processes of measurement, as when the unpredictable complexity of the particle showers are first trapped in the grids of detectors and then registered in the LED arrays in the lab. The chaos of the flashes in Roth's apparatus is placed in a subtle dialogue with both the phenomenon itself and the manner of its recording.

The ancient followers of Pythagoras characterized the harmonics of cosmic energy as the "music of the spheres." The mathematical foundations lay in the proportional system of the musical scale and in the perfections of Euclidean geometry. Here, by contrast, we stand as witnesses to the chaotic drumbeats of cosmic radiation. The new music is that of quantum mechanics and complexity — probabilistic rather than deterministic. A new art is encoding a new science.

Martin Kemp is emeritus professor in history of art at the University of Oxford, Oxford, UK.

See also page 847 and www.nature.com/astro09.

M. BREIG/FORSCHUNGSZENTRUM KARLSRUHE

Q&A: Tom Wolfe on language and the mind

Behind the novelist's eye of **Tom Wolfe** — bestselling author of *Bonfire of the Vanities* — lies a keen interest in brain science. Discussing the origin of language this week with Steven Pinker at the *Brainwave* festival in New York, Wolfe explains why he sees human behaviour as more than mechanistic, and genetic theory as little more than literature.

Your father was an agronomist. Did he give you a taste for science?

Actually, no. He worked at the Virginia Agricultural Experiment Station, dealing mainly in corn [maize], whose yield was enormously increased by artificial selection. My father edited a farm magazine called *The Southern Planter*. I would see him writing on a yellow legal pad and two weeks later his words were converted into beautiful, sharp-edged type. I was five years old, and it seemed magical. I lost track of the science. I just wanted to write.

When did you become interested in the brain?

When I enrolled in the American Studies graduate programme at Yale University I fell in love with sociology. I realized that almost everything one does is determined by status considerations — just think of your conduct in the bathroom, all by yourself. As early as 1956, I became convinced that there must be some area of the brain that determines this drive. I started trying to study brain physiology, only to discover that the status of Sigmund Freud's psychology was so high that the science had practically come to a halt. When it turned out that Freud did not have the answers — a fact brought out by lithium — only then did neuroscience, as we know it, get started.

Where did you begin your study?

The first thing I found was the book *Physical Control of the Mind* by the Spanish physician José Delgado at Yale. He was one of the first to insert needles into the brain to show where the seat of certain impulses were. Delgado proved his theory by letting himself be charged by a raging bull into whose brain he had sunk a stereotaxic needle. He pressed a button on a radio transmitter and the bull came to a screeching halt, like in the cartoons. He is still, to my thinking, the greatest neuroscientist ever.

Why are you sceptical about genetic explanations for human behaviour?

So many neuroscientists have become gnostics — convinced they see things the rest of us can't see because they've had a revelation. They have a secret: what we call



Tom Wolfe's interest in the brain was sparked by the mind-control experiments of José Delgado.

'soul', 'mind' and even 'self' must go into quotation marks. There's no 'me' inside of me. We are machines programmed at birth; we think we have free choice but we don't. But none of this has any scientific basis. As Delgado's son has said, we are not two miles down the long road of understanding the brain, we are two millimetres, and all the rest is literature.

What is your view of the origins of language?

It's my contention that evolution stopped when the first caveman spoke. Nobody knows when or how language was invented. Steven Pinker calls it an instinct that grows out of the process of evolution. My view is that language is an artefact like an axe or a sword, which can affect people's behaviour thousands of years later. Right now, as we speak, I'll bet you there are at least 250 million orgasms taking place that would not have taken place had Freud never lived. This is the influence of words. They will not lie down.

A *Forbes* article you wrote in 1996 made some think you wanted to defend the soul against the onslaught of science.

That is ludicrous. I don't care one way or another about the existence of the soul. I wrote the piece 'Sorry, but your soul just died' to point out that the march of neuroscience and genetic theory was beginning to prove Friedrich Nietzsche's prediction that the twenty-first century would see the eclipse of all values. If you have the gnostic belief that we are all just machines, and we react to one another according to how we are programmed, there is no room for values. I didn't say the theory was wrong. I predicted that, if it was proven correct, it would depress people.

Do you believe such a deterministic view of the brain could be validated?

My view in 1996 was, if they actually do end up proving it, I'd like to be there to report on it. It's all great journalism as far as I'm concerned — that's my sole interest. I now see that I made a mistake in trying to conflate neuroscience, which really is a science, with genetic theory, which is pure literature.

You predicted that, by 2006, brain imaging would be more important than the Internet. How has that held up?

Not very well. It's so easy to make predictions. Brain imaging has been very valuable in medicine, but I don't know of any great breakthroughs that have resulted. So much of it traces only blood flow. You bring up a certain subject to someone who's got the skull cap on and a certain part of the brain lights up, and you want to conclude that's where that subject is handled in the brain. But that's reading shadows on the wall. No, it hasn't done all those wonderful things. But eventually, in the hands of rational people, it will have tremendous promise. That's science, not literature. ■

Interview by **Jascha Hoffman**, a writer based in New York.

e-mail: jascha@jaschahoffman.com

Tom Wolfe's fourth novel is due out later this year. *Brainwave* events run until 23 April at the Rubin Museum of Art, New York.

NEWS & VIEWS

OBESITY

Be cool, lose weight

Stephen R. Farmer

To lose weight, would you rather diet, exercise or subject yourself to cool temperatures? The last choice is not such an odd one, as adult humans have brown fat tissue that burns calories in response to cold.

Fat is mainly stored in two types of adipose tissue: white and brown. White adipose tissue stores calories as large lipid droplets within individual cells. By contrast, brown adipose tissue (BAT) stores little fat, instead burning it to produce heat and regulate body temperature. Small mammals and newborn humans have copious amounts of BAT around their shoulder blades, which helps them to survive cold temperatures. Adult humans, however, were largely thought to lack BAT, with only one investigation¹ describing adipose tissue that seemed to function as BAT. Three independent studies^{2–4}, just published in *The New England Journal of Medicine*, follow up this observation, and conclusively identify BAT in adult humans.

BAT's ability to burn rather than store calories depends on each brown fat cell having many mitochondria — organelles that function as cells' power plants. Mitochondria are also abundant in the skeletal and heart muscles, and in the brain, where they convert metabolized sugars (glucose) and fats into the highly energized ATP molecule to fuel organismal activities. The mitochondria of brown fat cells are unique in that they contain UCP1, a protein that uncouples metabolism from ATP production in order to produce heat⁵.

To detect BAT in adult humans, all three studies^{2–4} used ¹⁸F-DG-based positron emission tomography/computerized tomography (PET/CT). This medical imaging technique measures the absorption of consumed ¹⁸F-fluorodeoxyglucose (¹⁸FDG) — a harmless radioactive form of glucose — into various tissues, providing information about the metabolic activity of each tissue. For definitive identification of BAT, the authors also performed histological and biochemical analysis of UCP1 in tissue biopsy samples.

In small mammals, exposure to cold stimulates the sympathetic nervous system to release the hormone adrenaline, which triggers brown fat cells to consume more fat and glucose for heat production. Virtanen *et al.*² scanned healthy volunteers on two separate days: on one day at normal room temperature; and on the other at 17–19°C while volunteers immersed one foot in cold (7–9°C) water

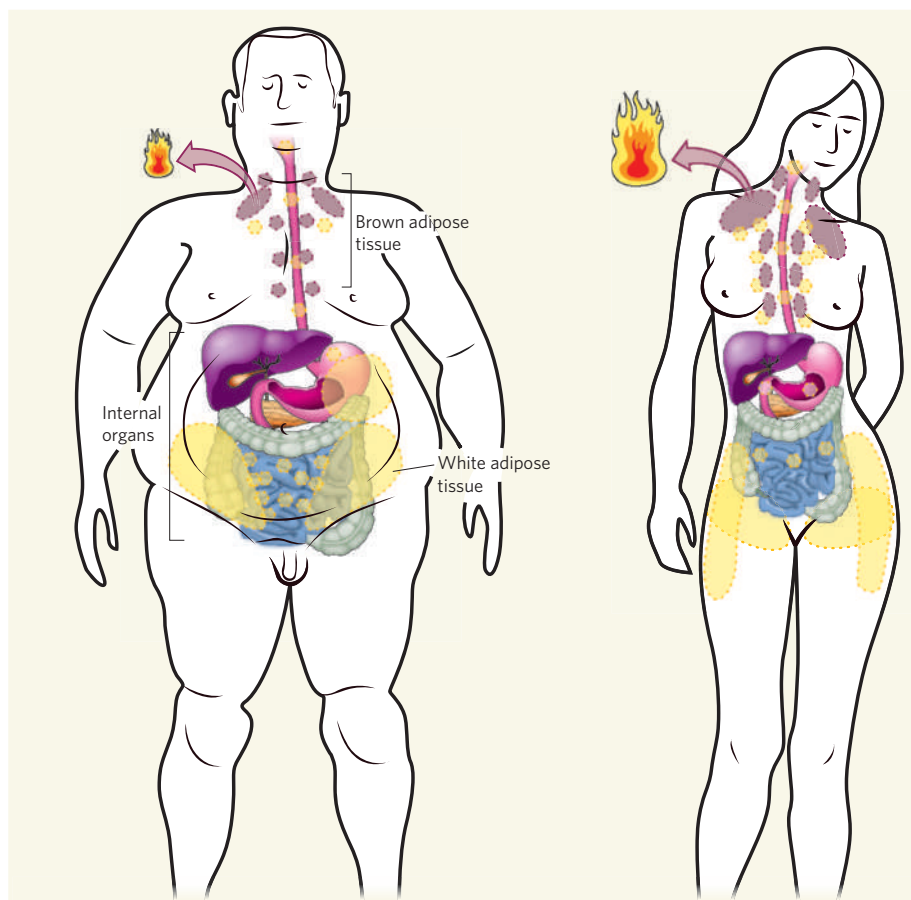


Figure 1 | When fat is good. Adult humans seem to contain brown adipose tissue (BAT) primarily behind the muscles of the lower neck and collarbone, as well as along the spine of the chest and abdomen^{2–4}. After food consumption, the absorbed fats and sugars are used to provide energy for daily functions, with excess calories being stored as fat in the white adipose tissue, which is mainly located under the skin of the buttocks and legs in women and around the internal organs in men. BAT can be activated in response to various stimuli, including exposure to cold, to burn fat and sugars. This process seems to be more prominent in the young and lean than in the old and obese, and in women rather than in men.

for 5 minutes, every other 5 minutes. In all individuals, exposure to cold led to a 15-fold increase in ¹⁸FDG uptake into the adipose tissue of the collarbone (supraclavicular) region. Tissue biopsy of precisely this region in three of the volunteers, based on morphology and UCP1 expression, revealed the presence of BAT. The authors detected 63 grams of supraclavicular BAT in one of the individuals — a mass they estimate could burn the equivalent

amount of energy during a year as is stored in about 4 kilograms of white fat tissue.

Van Marken Lichtenbelt *et al.*³ examined the presence, distribution and activity of BAT in 10 lean and 14 overweight/obese men in relation to body composition — percentage fat and lean-muscle mass — and energy metabolism. They report BAT activity in 23 of 24 individuals when measured at 16°C, but not at 22°C. The one individual with no BAT activity also

had the highest percentage (42%) of body fat. In fact, BAT activity within the group showed a significant negative correlation with percentage body fat and correlated positively with resting metabolic rate.

Excess storage of fat disrupts metabolic balance, leading to obesity-related disorders, such as diabetes and cardiovascular disease, which are collectively known as metabolic syndrome⁶. When fat accumulates in the intra-abdominal regions around internal organs — as it does mostly in men — the risk of developing metabolic syndrome is highest; women, by contrast, are at a lower risk, because they often store fat under the skin, around the thighs and buttocks⁷ (Fig. 1). Cypess and colleagues⁴ analysed 3,640 previously recorded scans of 1,972 patients (both men and women) who had undergone ¹⁸F-DG-PET/CT diagnostic screening for various medical reasons. The authors find scans corresponding to BAT activity in 7.5% of women and 3.1% of men. Moreover, women had a greater BAT mass, which absorbed more ¹⁸F-DG.

Larger amounts of active BAT also showed a positive correlation with younger age and lower outdoor temperature on the day each patient was scanned⁴. The authors detected an inverse relationship between active BAT and both smoking and patients' use of beta-blocker drugs for the treatment of high blood pressure and cardiovascular disease.

It is likely that the percentage of positive BAT scans Cypess *et al.* observed is lower compared with the other two reports^{2,3} because the authors⁴ relied on scans that were performed only at room temperatures. These studies, however, reach several similar conclusions: irrespective of age and gender, adult humans contain metabolically active BAT in their neck and upper chest regions; cold temperatures can activate BAT in adult humans, apparently more often in women than in men; and the presence of BAT correlates inversely with body fat, especially in older people.

Having reached pandemic levels worldwide, obesity and its related diseases have drastically increased health-care costs. With a role in adult-human metabolism, could BAT be exploited as a therapy for obesity?

For individuals with metabolically active BAT, one way to lose weight might simply be exposure to cold. As for others, years of investigation into the formation and function of BAT in mice has provided a wealth of knowledge that could aid the development of anti-obesity therapies targeting BAT in adult humans. For instance, weight loss might be achieved through drugs that mimic the cold by activating the sympathetic nervous system. Enhancing the formation of BAT, rather than white adipose tissue, might be another useful strategy. Such an enterprise has recently come closer to realization with the discovery^{8–10} of the stem-cell origins of the two adipose tissues. Moreover, the BMP7 protein has been identified¹¹ as the factor that selectively controls the

growth and development of brown fat cells, and so drugs that mimic its action might also be effective anti-obesity agents. ■

Stephen R. Farmer is in the Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118, USA. e-mail: farmer@biochem.bumc.bu.edu

1. Nedergaard, J., Bengtsson, T. & Cannon, B. *Am. J. Physiol. Endocrinol. Metab.* **293**, E444–E452 (2007).
2. Virtanen, K. A. *et al.* *N. Eng. J. Med.* **360**, 1518–1525 (2009).

3. van Marken Lichtenbelt, W. D. *et al.* *N. Eng. J. Med.* **360**, 1500–1508 (2009).
4. Cypess, A. M. *et al.* *N. Eng. J. Med.* **360**, 1509–1517 (2009).
5. Cannon, B. & Nedergaard, J. *Physiol. Rev.* **84**, 277–359 (2004).
6. Lean, M. E. J. *Proc. Nutr. Soc.* **59**, 331–336 (2000).
7. Wajchenberg, B. L. *Endocr. Rev.* **21**, 697–738 (2000).
8. Seale, P. *et al.* *Nature* **454**, 961–967 (2008).
9. Tang, W. *et al.* *Science* **322**, 583–586 (2008).
10. Rodeheffer, M. S., Birsoy, K. & Friedman, J. M. *Cell* **135**, 240–249 (2008).
11. Tseng, Y.-H. *et al.* *Nature* **454**, 1000–1004 (2008).

BIOCHEMISTRY

Anchors away

Maria Paola Costi and Stefania Ferrari

Nature often adopts several approaches to crack the same problem. The finding that the mechanism of a crucial enzyme in certain disease-causing bacteria differs from that in mammals offers scope for drug discovery.

On page 919 of this issue, Koehn *et al.*¹ propose that, in certain microorganisms, a previously unknown biochemical mechanism underpins the function of an enzyme that is essential to the microorganisms' survival. In mammals, the activity of this enzyme — thymidylate synthase — depends on an 'anchor' in its active site that binds covalently to the enzyme's substrate. But the authors find that, in some microbes (including many that threaten human life), thymidylate synthase is active in the absence of such an anchor. The mechanism that explains this behaviour is a potential target for antibiotic drugs that would be toxic to microbes, but not to humans.

This difference¹ between taxonomic groups is a clear example of how some cells evolved to have well-developed enzyme mechanisms that have high energy costs (as in mammalian thymidylate synthase), whereas others make do with less-specialized mechanisms that have lower energy costs (as in the microbial enzyme). Thymidylate synthase produces a deoxythymidine nucleotide (dTMP), which is necessary for DNA synthesis. Classic biochemical² and proteomic studies³ have clearly shown that there are two kinds of thymidylate synthase, each having distinct evolutionary origins (based on their different mechanisms of action and structures). Those found in humans and other mammals are known as TS enzymes, whereas the other group, found in 30% of microbial genomes⁴, is known as the FDTS family of enzymes.

The mechanistic differences between the two groups hinge on the cofactors required and on the reactions that occur between the enzymes and their substrate, a deoxyuridine nucleotide (dUMP). In mammalian TS, synthesis of dTMP begins when a cysteine amino-acid residue in TS forms a covalent bond to a specific carbon in dUMP (see Fig. 1 here, and Fig. 1a on page 920). This bond anchors the

substrate to the enzyme. In the next step, a carbon atom is transferred from a cofactor to the carbon adjacent to the anchor. This is a high-energy process, which occurs only because the anchor aligns the reacting groups perfectly for reaction. In the final step, the newly attached carbon atom is converted into a methyl group and the anchor breaks, releasing the resulting dTMP product.

But Koehn *et al.*¹ have found that microbial FDTS enzymes do not covalently anchor dUMP, reducing the energy cost of their reactions. To prove this, they performed conceptually simple experiments on the active site of FDTS from the microbe *Thermotoga maritima*. It had previously been thought that a serine amino-acid residue in the active site acts as an anchor for dUMP, in the same way as a cysteine residue does in TS enzymes. The authors therefore mutated the serine to alanine, whose side chain is incapable of reacting with dUMP to form a covalent anchor. The mutant FDTS retained its activity, thus showing that substrate anchoring is not necessary to drive the enzyme's reaction.

The authors also made a serine-to-cysteine mutant, which was expected to be active by analogy with the TS enzymes. In fact, it was less active than the non-mutated enzyme. A crystal structure¹ of the mutant revealed that the cysteine residue does not form a covalent bond to dUMP, yet Koehn *et al.* obtained other evidence suggesting that the cysteine does form a complex with the substrate. Taken together, these results suggest that the observed cysteine–dUMP complex is a dead end that doesn't form part of the FDTS catalytic cycle. These data serve as a reminder that, although mutagenesis experiments are often very useful, local changes to protein structures can translate into mostly unpredictable long-range effects. Indeed, previous mutagenesis experiments^{4,5} on the FDTS of the bacterium *Helicobacter pylori*

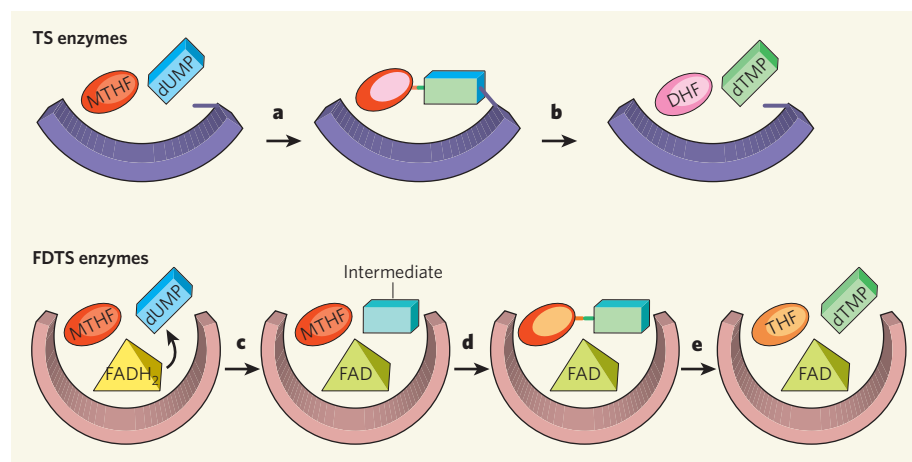


Figure 1 | Different mechanisms in the biosynthesis of dTMP. Thymidylate synthase enzymes make deoxythymidine nucleotide (dTMP), which is necessary for the synthesis of DNA. Two groups of these enzymes exist: TS enzymes in mammals and FdTS enzymes in certain microorganisms. **a**, In TS enzymes, the substrate (deoxyuridine nucleotide, dUMP) is first anchored to the active site, whereupon it reacts with a cofactor (MTHF). **b**, The products of the reaction (dTMP and the spent cofactor, DHF) are then released. **c**, Koehn *et al.*¹ propose that, in FdTS enzymes, dUMP first reacts with a different cofactor, FADH₂, to form an intermediate and a side product (FAD). **d**, **e**, The intermediate reacts with MTHF to form dTMP and another side product, THF. No anchoring of dUMP occurs in FdTS enzymes.

produced misleading evidence about the mechanism of action of the enzymes.

The reactivity of the dUMP molecule (specifically, of its uracil nucleotide base) in the FdTS active site is very different from its reactivity in the TS active site, providing another indication of differences between the two classes of enzymes. Koehn *et al.*¹ propose that a third component of FdTS reactions, the 'FADH₂ cofactor', is responsible for this difference. They claim that an unusual reaction occurs in which a hydrogen atom is transferred from FADH₂ to a carbon in the uracil base — the same carbon that is anchored covalently in TS enzymes (see Fig. 1c on page 920).

The authors found evidence for this theory by introducing hydrogen isotopes into the various components of the FdTS reaction,

and following the movements of the isotopes as the reaction progressed. They thus showed that FADH₂ is the source of the hydrogen atom that ends up in the uracil core of dUMP. Further evidence came from the authors' X-ray crystal structures of FdTS in complex with analogues of dUMP and with FAD (the side product formed from FADH₂). The distance between the active-site serine and the uracil carbon atom is too great to allow a covalent bond to form between them (as occurs in TS enzymes). Instead, the FAD molecule lies close to the reactive carbon atom, as would be expected if the FADH₂ cofactor transfers a hydrogen atom to dUMP's uracil. The complete picture emerging from Koehn and colleagues' multidisciplinary work convincingly shows that the reaction pathway occurring in certain

microbial thymidylate synthases is different from that occurring in mammals.

The unravelling of the entire FdTS pathway through to the end products remains incomplete, and other loose ends must still be tied up — for example, the formation of the 'dead-end complex'. As it is unlikely that FdTS is simply a reaction vessel for dUMP and the two cofactors, the enzyme's contribution to the mechanism must also be clearly characterized in the future.

Although Koehn and colleagues work¹ raises the possibility of developing antibiotics that selectively block microbial FdTS, but not mammalian TS, this remains to be demonstrated, and obstacles certainly exist: only a few inhibitors⁶ of FdTS have been discovered since the protein was first reported in 2002. Nevertheless, the prospect of such an antibiotic is exciting, as several pathogenic bacteria, including that responsible for tuberculosis, should be susceptible. More broadly, the authors have discovered a new concept in antibiotic drug discovery — compounds that resemble an enzyme's substrate, but that incorporate significant changes to the core structure (such as that occurring in dUMP when it accepts a hydrogen atom from FADH₂), might selectively attack bacteria, and so be less toxic to humans.

Maria Paola Costi and Stefania Ferrari are in the Department of Pharmaceutical Science, Università degli Studi di Modena e Reggio Emilia, via Campi 183, Modena 41100, Italy. e-mails: costimp@unimore.it; stefania.ferrari@unimore.it

1. Koehn, E. M. *et al.* *Nature* **458**, 919–923 (2009).
2. Myllykallio, H. *et al.* *Science* **297**, 105–107 (2002).
3. Lesley, S. A. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 11664–11669 (2002).
4. Leduc, D. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 7252–7257 (2004).
5. Agrawal, N., Lesley, S. A., Kuhn, P. & Kohen, A. *Biochemistry* **43**, 10295–10301 (2004).
6. Esra Oenen, F. *et al.* *Bioorg. Med. Chem. Lett.* **18**, 3628–3631 (2008).

SOLID-STATE PHYSICS

Bouncing spins

Lieven M. K. Vandersypen

The conventional approach to flipping electron spins in a semiconductor requires an external alternating field. It seems that the same job can be accomplished without external excitation of any kind.

A mother pushing her son on a swing knows that she must push the swing in synchrony with the swinging motion. Similarly, to flip the spin of an electron in a static magnetic field, we must apply an alternating magnetic field whose frequency matches the wobbling motion of the electron's spin orientation about the static field. Or so we thought. In their report on page 868 of this issue, Frolov *et al.*¹ show that

electron spins can be flipped with electric fields, which is unusual in itself, and that the fields are purely static. This sounds as if the mother were pushing the swing in the same direction all the time, but the swing nevertheless rocks back and forth.

Spin flips have been achieved using external electric fields in previous studies, but these fields were always oscillating, such that electrons

in the material under study were pulled back and forth periodically. This in turn, by means of a quantum effect known as spin–orbit interaction, creates the alternating magnetic field required to flip the spin: when an electron moves through the material, its spin experiences a magnetic field that is proportional to the electron's velocity (orbit). The orientation of this field is reversed when the electron's direction of motion is reversed. This approach has been used to flip spins in a variety of systems, from two-dimensional gases of electrons² to 'quantum dots'³. But what would be desirable for electronic applications is the ability to manipulate spins without any external excitation, using for example static electric fields, which are easier to generate on a chip than oscillating fields.

Using a clever and original approach, Frolov *et al.*¹ accomplish just that by making electrons

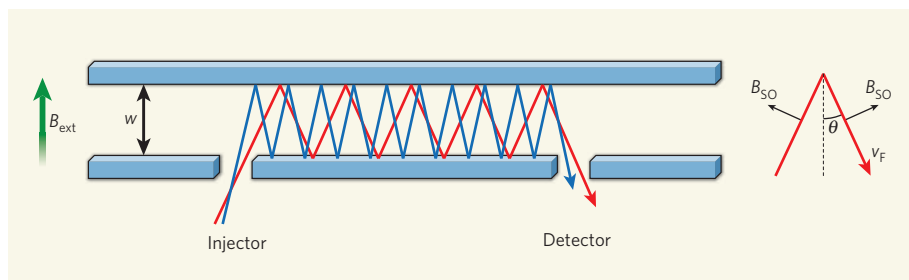


Figure 1 | Flipping electron spins. Electrons injected into a semiconductor channel through a narrow opening (injector), which filters electrons with spins oriented along a static magnetic field (B_{ext}), bounce back and forth between the channel walls and experience an alternating magnetic field (B_{SO}) through a process known as spin–orbit interaction. A second opening (detector), 20 micrometres down the channel, verifies whether electron spins have been flipped by the oscillating field. Frolov *et al.*¹ show that if the bouncing frequency (which depends on the channel width, w , the characteristic velocity, v_F , with which electrons move through the channel, and the deflection angle, θ) matches the frequency of the alternating magnetic field, electron spins will be flipped.

bounce back and forth between the two walls of a micrometre-wide semiconductor channel using solely static electric fields. The spin–orbit-induced magnetic field then alternates with a frequency that corresponds to the electron's bouncing frequency, which depends only on the separation between the channel walls and on the Fermi velocity — the characteristic velocity with which electrons move through the semiconductor.

In the authors' experiment, electrons are injected into the semiconductor channel via a narrow opening that filters electrons, selecting those with a spin axis oriented along a static magnetic field. The electrons then move through the channel, bouncing back and forth between the channel walls along the way. A second opening, 20 micrometres down the channel, acts as a spin detector and thus tests whether the electron spins have been flipped or not (Fig. 1).

But how do Frolov and colleagues know whether their technique of flipping electron spins really works? After all, there can be other reasons — apart from the bouncing between channel walls — for electron spins to flip between injector and detector⁴. To confirm their technique, the authors performed two tests. The first test consisted of assessing whether a spin flip occurred only when the electron's bouncing frequency matched the spin resonance frequency (the magnetic-field frequency necessary to trigger the spin flip), which is proportional to the strength of the static magnetic field. And indeed, spin flips were detected only for a specific magnetic field strength. Furthermore, the required magnetic-field strength changed as predicted with the channel width, as well as with the Fermi velocity (which is controlled by the electron density), both of which affect the bouncing frequency.

The second crucial test involved aligning the static magnetic field along the channel axis rather than perpendicular to it, as was the case in the first test. The spin–orbit-induced magnetic field was then parallel to the induced field expected, which should not — and did not — lead to spin flips. The two tests thus

confirm the authors' novel technique as a robust method of flipping spins.

But there's a catch. As far as electronic applications are concerned, the technique comes with a major limitation: it does not flip electron spins in a coherent manner, but rather does so randomly — that is, it cannot rotate the electron spins to a specific orientation. Although the authors used an ultraclean semiconductor material, electrons still scatter off charged impurities in the vicinity of the channel, randomizing the bouncing path, and hence the rotation of their spins. In addition, electrons enter the semiconductor channel and hit the

walls with a range of incidence angles (Fig. 1), leading to a large array of bouncing frequencies and hence of spin orientations. But it may be possible to circumvent this limitation by using even cleaner materials and, as Frolov and colleagues suggest, by applying electron-focusing techniques to obtain a well-defined angle of incidence and a more sharply defined bouncing frequency.

In the future, one could imagine applying the authors' technique of flipping spins to entire solid-state electronic circuits in which information is encoded in the spin state of the electrons. This is the vision of the field of spintronics⁵, which has already led to discoveries such as 'giant magnetoresistance' and the subsequent miniaturization of hard-disk drives. One step beyond lies the coherent control of quantum-mechanical superpositions of electron spin states⁶, which may one day lead to applications in quantum-information processing.

Lieven M. K. Vandersypen is at the Kavli Institute of NanoScience, TU Delft, Lorentzweg 1, 2628 CJ Delft, the Netherlands.

e-mail: l.m.k.vandersypen@tudelft.nl

1. Frolov, S. M. *et al.* *Nature* **458**, 868–871 (2009).
2. Kato, Y., Myers, R. C., Gossard, A. C. & Awschalom, D. D. *Nature* **427**, 50–53 (2004).
3. Nowack, K. C., Koppens, F. H. L., Nazarov, Yu. V. & Vandersypen, L. M. K. *Science* **318**, 1430–1433 (2007).
4. Meier, F. & Zakharchenya, B. P. (eds) *Optical Orientation* (North-Holland, 1984).
5. Wolf, S. A. *et al.* *Science* **294**, 1488–1495 (2001).
6. Hanson, R. *et al.* *Rev. Mod. Phys.* **79**, 1217–1265 (2007).

NEUROSCIENCE

A bar code for differentiation

Nicholas C. Spitzer

Regulating neuronal development can be complicated. But genetic control of neurotransmitter expression — as exemplified by differentiation of dopamine-secreting neurons — turns out to be relatively straightforward.

Specifying how nerve cells differentiate is a Herculean challenge: a huge number of neurons with extraordinary diversity of elaborate architectures and sophisticated functions materialize almost magically during a brief period of embryonic development. Differentiation of neurons into groups with distinct characteristics is known to require coordinated synthesis of many proteins, including enzymes that produce, and transporters that package, neurotransmitters. Success results from the orchestrated expression of batteries of genes, but the way in which this is achieved has remained unclear. In a remarkable paper in this issue (page 885), Flames and Hobert¹ demonstrate how the differentiation of neurons of different lineages that share the ability to secrete the neurotransmitter dopamine is regulated at the level of gene transcription.

Because cells with common features often

originate from distinct cell lineages, different populations of cells could acquire common characteristics when different switches turn on the same genes. To investigate whether this is the case for dopamine-secreting (dopaminergic) neurons, Flames and Hobert analysed neuronal development in the nematode worm *Caenorhabditis elegans* — a relatively simple organism with a nervous system of 302 neurons, which include four classes of lineage-unrelated dopaminergic nerve cells. They found that the five genes that encode proteins for the synthesis and transport of dopamine (the dopamine-pathway genes) share an evolutionarily conserved DNA sequence in their regulatory regions. This sequence was sufficient to drive the expression of these genes in dopaminergic neurons, and its mutation abolished their expression.

On further inspection, the authors narrowed

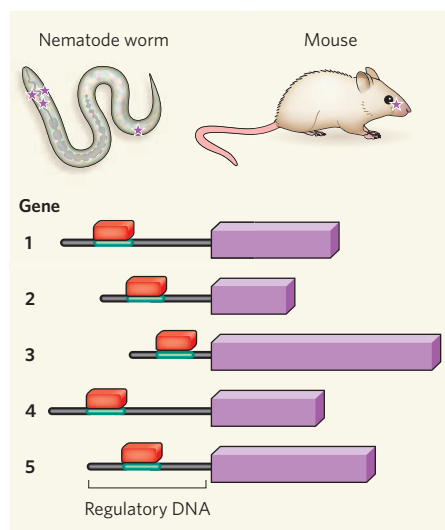


Figure 1 | Orchestrating neurotransmitter expression. Flames and Hobert¹ show that, in neurons of both the nematode worm and the mouse, binding of the same transcription factor (red) to the same short DNA bar-code motif (green) coordinates activation of all five genes (purple) required for expression of the neurotransmitter dopamine. Stars indicate the locations of neurons examined in each species.

down the sequence to a shorter one characteristic of the binding site for the ETS class of transcription factors, and named it the dopamine motif. Of the various ETS proteins they analysed, AST-1 seems to be both necessary and sufficient to activate concerted expression of dopamine-pathway genes by binding to the dopamine motif during development. Moreover, this transcription factor is required for sustained expression of dopamine in mature neurons¹ — an insight that is particularly relevant for a class of neurons the degeneration of which causes Parkinson's disease. So it seems that a simple 'bar code' — in the form of the dopamine motif — specifically and reliably coordinates the expression of genes that drive the differentiation of dopaminergic neurons.

Flames and Hobert¹ extend their findings to mice, thus showing that the regulatory mechanism they describe is evolutionarily conserved. They find that the ETS transcription factor ETV1 is present in dopaminergic neurons in the mouse olfactory bulb. Blocking the expression of ETV1 reduced the number of these neurons, and enhancing its expression increased the number, through the same conserved dopamine motif found in all five mouse dopamine-pathway genes (Fig. 1). What's more, mouse ETV1 could overcome the effects of *ast-1* mutations in the worm, further supporting the conserved function of ETS transcription factors.

Is the mechanism that specifies differentiation of dopaminergic neurons unique? Previous work on similar problems in *C. elegans* set the stage for Flames and Hobert's study. The specification of a group of six mechanosensitive neurons is, for example, governed by the action of two transcription factors that bind to a DNA

motif necessary for the development of touch sensitivity². And the identities of the ASE class of sensory neurons and AIY class of interneurons in *C. elegans* are specified by comparable motifs^{3,4}. In addition, this general strategy for regulation of gene expression is not restricted to *C. elegans*: in the vertebrate retina, similar regulatory organization determines the fate of photoreceptor neurons through a handful of DNA motifs spread among numerous genes expressed in these cells⁵. What is unique about Flames and Hobert's observations¹, however, is that the mechanism they describe crosses neuronal lineages and spans species, providing cogent evidence for its generality.

It is well established that gene activation can require the binding of combinations of transcription factors to DNA⁶. Does such combinatorial coding mediate differentiation of dopaminergic neurons? Flames and Hobert find that expression of AST-1 in neurons that do not normally express it drives activation of dopamine-pathway genes in some but not all neurons, and only at particular times during development. This result indicates that AST-1 action depends on specific molecular and temporal contexts — a feature reminiscent of the context-dependent expression of 'master genes' such as the *myoD* gene, which can drive muscle development⁷, and the *eyeless* gene, which can drive expression of ectopic eyes⁸. Context dependence generally indicates combinatorial coding, and so coding partners for AST-1 are likely to be identified.

Flames and Hobert's data¹ suggest that the conserved dopamine DNA motif and the transcription factor that activates it are key to regulating the expression of this neurotransmitter. More generally, their results support a scheme⁹ whereby genes that influence the terminal differentiation of cells — known as terminal selector genes — encode transcription factors that bind to simple, common DNA-selector motifs on cohorts of genes that determine specific neuronal properties. Sherlock Holmes' bemused delight in breaking the Secret Code is surely matched by the satisfaction provided by this compelling demonstration of a shared DNA bar code underlying a crucial aspect of neuronal differentiation. ■

Nicholas C. Spitzer is in the Neurobiology Section, Division of Biological Sciences and Center for Molecular Genetics, Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, California 92093-0357, USA. e-mail: nspitzer@ucsd.edu

1. Flames, N. & Hobert, O. *Nature* **458**, 885–889 (2009).
2. Duggan, A., Ma, C. & Chalfie, M. *Development* **125**, 4107–4119 (1998).
3. Etchberger, J. F. et al. *Genes Dev.* **21**, 1653–1674 (2007).
4. Wenick, A. S. & Hobert, O. *Dev. Cell* **6**, 757–770 (2004).
5. Hsiao, T. H.-C. et al. *PLoS ONE* **2**, e643 (2007).
6. Shirasaki, R. & Pfaff, S. L. *Annu. Rev. Neurosci.* **25**, 251–281 (2002).
7. Weintraub, H. et al. *Proc. Natl Acad. Sci. USA* **86**, 5434–5438 (1989).
8. Halder, G., Callaerts, P. & Gehring, W. J. *Science* **267**, 1788–1792 (1995).
9. Hobert, O. *Proc. Natl Acad. Sci. USA* **105**, 20067–20071 (2008).

IMAGING

Nanoscale MRI

P. C. Hammel

Magnetic resonance imaging offers rich three-dimensional pictures, but with limited resolution. Imaging at the nanometre scale has now become possible using highly sensitive force-detection techniques.

Seeing inside a complex object is an invaluable aid to understanding it, so three-dimensional imaging is a pressing objective in areas ranging from molecular and cell biology to investigations of the electronic and structural properties of materials. The challenge is made more difficult by the desire to see the object without altering it in the process. This requires a delicate touch, involving as weak an interaction as possible with the object. But such an approach often conflicts with the need for the high spatial resolution that makes fine detail visible.

Writing in *Proceedings of the National Academy of Sciences*¹, a group led by Dan Rugar reports the success of a delicate yet effective approach to imaging — one that gently excites the nuclear spins in their test objects, particles of tobacco mosaic virus, and records their locations by listening to the spins' oscillating mag-

netization. An image is derived by recording these signals from a three-dimensional mesh of locations within the object; achieving high resolution requires the mesh to be very fine, so that the volume sampled at each grid location is as small as possible. The authors' sensitive detection of the feeble signals from these elements of nanometre-scale volume, and subsequent reconstruction of the three-dimensional structure of the virus from these signals, marks the arrival of a powerful tool for non-perturbative imaging of a single copy of an object — be it biological, electronic or magnetic — at the nanometre scale.

Rugar and colleagues' approach is inspired by the impressive three-dimensional views of a material provided by magnetic resonance imaging (MRI). Rather than measuring how energetic particles interact with the object

to obtain an image, the technique uses radio waves whose energy is less than a billionth of that of the X-rays used for diffraction studies or the electrons used in an electron microscope. MRI is itself based on nuclear magnetic resonance (NMR), which exploits the intrinsic and plentiful nuclear magnetic spins present in all substances. These nuclear magnets oscillate at a precisely measurable frequency that is determined by fields generated by neighbouring atoms, and by an externally applied field. Hence, these nuclear magnets are embedded, microscopic probes that reveal details of their host's electronic, magnetic and structural properties. Detailed information obtained from NMR has been extensively used for tasks ranging from identifying organic molecules to illuminating subtle features of exotic superconductors.

For imaging, the external field is arranged to vary controllably across the sample, so that the frequency of the nuclear magnetic oscillation will reveal its precise location. This mechanism underlies non-invasive, three-dimensional MRI of regions deep within a sample. Rather than scattering energetic particles, MRI uses low-energy radio waves to excite the nuclear spins so that their oscillation frequency can be measured. A benefit of using magnetic resonance for imaging is that these magnetic resonance signals allow spatially resolved NMR experiments and characterization that enrich the images with detailed microscopic information.

However, the weak interaction that makes MRI so non-invasive is also its Achilles heel: the interaction of the detector with the spin is so small that, in conventional approaches, many spins (10^{12} – 10^{18}) are needed to provide a large enough signal to tease out information about the materials. The dimensions of the resolvable volume are limited by the need to detect the weak oscillatory signal of the few spins in the small-volume elements that make up the image. This limits conventional MRI to volumes of several cubic micrometres, and so reduces the usefulness of the technique in solid-state physics, or molecular or cell biology.

In 1991, John Sidles² proposed a system for mechanically sensing the weak force that a microscopic ferromagnet exerts on the nuclear magnetic moment in a sample. Tiny forces, he suggested, can be measured by placing the sample under investigation on a compliant cantilever. By observing the slight resulting deflection of the cantilever using, for example, an optical interferometer, extraordinarily small forces can be detected³. Force-detected MRI, dubbed magnetic resonance force microscopy (MRFM), has rapidly improved in sensitivity and spatial resolution^{4–6}: it has been used to observe a single electron spin⁷ and for highly sensitive nuclear-spin detection⁸. MRFM is also a practical materials probe that has been applied to major problems in science^{9,10} and technology¹¹. Beyond this, it has been shown that techniques used in conventional pulsed NMR are effective for force-detected magnetic resonance¹².

Rugar and colleagues' imaging of individual

virus particles¹ is a striking advance in MRFM capability that demanded exceptional detection sensitivity. In particular, the ferromagnetic probe must be brought within tens of nanometres of the cantilever-mounted virus. At these distances, the cantilever experiences many other forces from the nearby surfaces — including, for example, van der Waals forces that are typically thousands to millions of times larger than the nuclear magnetic forces to be measured, and dissipative, electrostatic cantilever-surface forces that produce noise that obscures the signal.

The authors' success is the fruit of a decade of work developing ultrasensitive force-detection techniques¹³, which manipulate the spins to produce a distinctive force signal that can be picked out from the background forces, and a nanofabricated antenna¹⁴ that produces a strong radio-frequency magnetic excitation field sufficiently localized that it doesn't disturb the cantilever (the nuclear magnetic forces generate cantilever deflections only at the sub-angstrom level). Finally, the work shows that the noisy signals can be deconvolved into images.

The MRFM procedure will not meet all imaging needs. It is a demanding technique that must be performed in a vacuum and at low temperature. This is a limitation that is shared by electron microscopy of biological specimens, which is nonetheless a highly successful imaging tool. The detection sensitivity of MRFM is improving rapidly, and its history indicates that these capabilities, now at the cutting edge, will soon be routine for MRFM practitioners. But it will be some time before those capabilities can be exploited

by the wider microscopy community.

That said, the demonstration¹ of the imaging of viral particles at a resolution down to 4 nanometres heralds the emergence of a new microscope for investigating native biological specimens that will compete with, and complement, electron microscopy and NMR spectroscopy. It uniquely combines non-destructive imaging with the capability of imaging individual copies of specimens such as proteins. The approach is also likely to find wide application beyond biology, in investigations of the chemical and elemental make-up of nanostructures in the physical and materials sciences. ■

P. C. Hammel is in the Department of Physics, Ohio State University, Columbus, Ohio 43210, USA.

e-mail: hammel@mps.ohio-state.edu

1. Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T. & Rugar, D. *Proc. Natl Acad. Sci. USA* **106**, 1313–1317 (2009).
2. Sidles, J. A. *Appl. Phys. Lett.* doi:10.1063/1.104757 (1991).
3. Mamin, H. J. & Rugar, D. *Appl. Phys. Lett.* doi:10.1063/1.1418256 (2001).
4. Sidles, J. A. *et al. Rev. Mod. Phys.* doi:10.1103/RevModPhys.67.249 (1995).
5. Hammel, P. C. & Pelekhov, D. V. *Handbook of Magnetism and Advanced Magnetic Materials* Vol. 5 (Wiley, 2007).
6. Kuehn, S., Hickman, S. A. & Marohn, J. A. *J. Chem. Phys.* doi:10.1063/1.2834737 (2008).
7. Rugar, D. *et al. Nature* **430**, 329–332 (2004).
8. Mamin, H. J., Poggio, M., Degen, C. L. & Rugar, D. *Nature Nanotechnol.* doi:10.1038/nnano.2007.105 (2007).
9. Obukhov, Y. *et al. Phys. Rev. Lett.* doi:10.1103/PhysRevLett.100.197601 (2008).
10. Klein, O. *Phys. Rev. B* doi:10.1103/PhysRevB.78.144410 (2008).
11. Thurber, K., Harrell, L. & Smith, D. J. *Mag. Res.* doi:10.1016/S1090-7807(03)00040-5 (2003).
12. Lin, Q. *et al. Phys. Rev. Lett.* doi:10.1103/PhysRevLett.96.137604 (2006).
13. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. *AIP Conf. Proc.* **696**, 45 (2003).
14. Poggio, M., Degen, C. L., Rettner, C. T., Mamin, H. J. & Rugar, D. *Appl. Phys. Lett.* doi:10.1063/1.2752536 (2007).

MATERIALS SCIENCE

Nanotubes unzipped

Mauricio Terrones

Nanotubes are single sheets of graphite rolled up into a cylinder. But no one thought that nanotubes could be cut along their axis and flattened out to make such sheets. Until now.

The discovery of buckyballs and carbon nanotubes in the 1980s and early 1990s^{1–3} launched the field of carbon nanoscience, and spawned intensive research into the synthesis and applications of these structures. For a long time, it seemed as if the landscape of the carbon nanoworld contained only round objects — spheres and tubes. But in the twenty-first century, flat forms of carbon gained prominence with the discovery of graphene⁴ (single layers of graphite) and graphene nanoribbons^{5,6}. To realize the practical potential of these newcomers, methods for their mass production are sorely needed. In this issue, two possible solutions are reported — by Kosynkin *et al.*⁷ (page 872) and Jiao *et al.*⁸ (page 877) — in

which nanotubes are 'unzipped' and rolled open to produce nanoribbons.

Graphene is a metal-like conductor, but nanoribbons can generally be either metallic or semiconducting depending on the patterns formed by their edges⁵. Furthermore, nanoribbons less than 10 nanometres wide are expected to be semiconductors, independent of their edge patterns. Narrow nanoribbons are thus excellent candidates for use in electronic devices, such as field-effect transistors, which form the basis of microchips in computers. A thorough exploration of the chemical and mechanical properties of nanoribbons will undoubtedly suggest other applications for these structures, perhaps as sensors, catalysts,

scaffolds for tissue regeneration or components of composite materials.

Existing methods for making nanoribbons involve chemical synthesis, cutting graphene sheets into ribbons, or using ultrasound to break up graphene that has had its surface modified by the non-covalent binding of polymer molecules. But these methods produce only minute quantities of nanoribbons. A technique for producing bulk quantities has been reported⁹, which involves depositing volatile carbon precursors onto a substrate where they react to form nanoribbons that are metal conductors. Nevertheless, alternatives to this chemical vapour deposition method still need to be developed that produce large-scale amounts of semiconducting nanoribbons.

Kosynkin *et al.*⁷ report an extremely simple, efficient and potentially scalable technique for making graphene sheets and nanoribbons. The authors' starting materials are multiwalled nanotubes consisting of 15–20 concentric cylinders, with diameters of 40–80 nanometres. The method involves treating the nanotubes with concentrated sulphuric acid followed by potassium permanganate (an oxidizing agent) at room temperature, and finally heating them at 55–70 °C (Fig. 1a). This process chemically unzips the nanotubes, forming nanoribbons up to 4 micrometres long, with widths of 100–500 nanometres and thicknesses of 1–30 graphene layers. The products are highly soluble in water and in polar organic solvents, which is crucial if the nanoribbons are to be used in composite materials or for biological applications.

The chemical mechanism of the unzipping process probably involves the oxidation of carbon–carbon double bonds in the nanotubes. But it could also be that sulphuric acid molecules insert themselves between the concentric cylinders of the nanotubes — a similar 'intercalation' occurs when graphite is treated with sulphuric acid and potassium permanganate to peel off graphene sheets. The mechanism of Kosynkin and colleagues' technique thus needs clarification, and should stimulate further experiments.

The authors found that their nanoribbons were poor conductors, because the edges of the structures hold many oxygen-containing chemical groups that disrupt the flow of charge carriers. Kosynkin *et al.* therefore removed these groups by treating their products with a reducing agent, or by heating (annealing) the products in hydrogen. The wide nanoribbons thus produced were metallic conductors, similar to those grown by chemical vapour deposition. The authors also showed that their chemically reduced nanoribbons are in principle suitable for making field-effect transistors. Another benefit of the annealing process is that it could improve the reactivity and smoothness of the nanoribbons' edges.

Kosynkin and colleagues also used their method to unzip single-walled carbon nanotubes to yield narrow nanoribbons. Unfortunately, the resulting products become

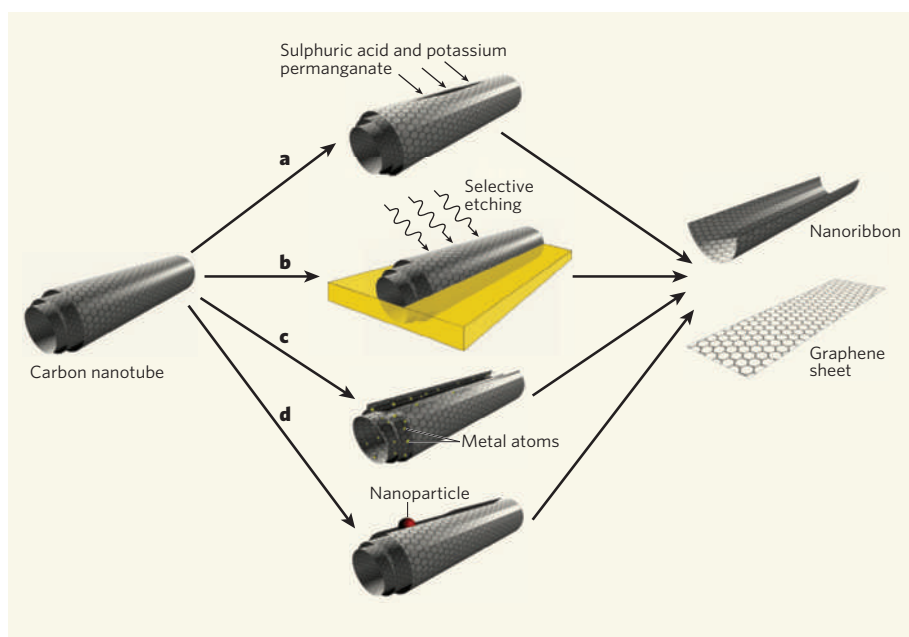


Figure 1 | Methods for unzipping carbon nanotubes. **a**, Kosynkin *et al.*⁷ report that multiwalled carbon nanotubes can be unzipped by treating them with sulphuric acid and potassium permanganate (an oxidizing agent) to form nanoribbons or graphene sheets (single layers of graphite). **b**, Jiao *et al.*⁸ describe a complementary method, in which nanotubes partially embedded in a polymer film are etched by argon plasma. **c**, Another approach¹⁰ is to insert alkali-metal atoms between the concentric cylinders of a multiwalled carbon nanotube, which causes graphene sheets to peel off. **d**, A method still to be explored would use catalytic metal nanoparticles to cut along the length of a nanotube like a pair of scissors. (Graphic by A. R. Botello-Méndez.)

entangled; further experiments are therefore being done to find ways of untangling the ribbons so that they can be of practical use.

The authors' technique works well with nanotubes that have many structural defects on their surfaces (such as those made by chemical vapour deposition). But it is less effective with more crystalline nanotubes produced by other methods, such as laser ablation or arc discharge. Fortunately, Jiao *et al.*⁸ describe an alternative approach for unzipping highly crystalline multiwalled carbon nanotubes. They partially embedded tubes in a polymer film, and then etched them with argon plasma (Fig. 1b). The film was then removed using solvent vapour, and the resulting nanoribbons were heated at 300 °C to remove any residual polymer.

The thicknesses of Jiao and colleagues' nanoribbons typically ranged from one to three graphene layers, depending on the plasma etching conditions. The ribbons were also narrower (10–20 nanometres wide) than those of Kosynkin *et al.*⁷. As expected, Jiao and colleagues' narrow ribbons⁸ were semiconductors (unlike Kosynkin and colleagues' wider ribbons, which were metallic conductors).

The two reports^{7,8} break new ground in the bulk fabrication of nanoribbons. An alternative method for unzipping multiwalled carbon nanotubes has also just been reported¹⁰, in which alkali-metal atoms intercalate between the concentric cylinders of the nanotubes. The atoms are then washed out, which causes the tubes to open along their axes (Fig. 1c).

Furthermore, catalytic particles of metals such as iron and nickel can cut through graphene sheets¹¹. This effect could also be used to unzip multiwalled carbon nanotubes to produce nanoribbons¹², and should be explored further (Fig. 1d).

More research is, however, needed to find ways of efficiently unwrapping single- and double-walled nanotubes, in order to carefully control the widths and edge patterns of nanoribbons. Once bulk quantities of nanoribbons are available, their toxicological effects and possible biological applications can be studied. And, last but not least, the potentially unusual magnetic and catalytic properties of these materials can finally be explored. ■

Mauricio Terrones is in the Laboratory for Nanoscience and Nanotechnology Research, and the Advanced Materials Department, Instituto Potosino de Investigación Científica y Tecnológica, San Luis Potosí 78216, Mexico. e-mail: mterrone@ipicyt.edu.mx

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
2. Endo, M. *Chemtech* **18**, 568–576 (1988).
3. Iijima, S. *Nature* **354**, 56–58 (1991).
4. Geim, A. K. & Novoselov, K. S. *Nature Mater.* **6**, 183–191 (2007).
5. Nakada, K., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. *Phys. Rev. B* **54**, 17954–17961 (1996).
6. Li, X. L. *et al.* *Science* **319**, 1229–1232 (2008).
7. Kosynkin, D. V. *et al.* *Nature* **458**, 872–876 (2009).
8. Jiao, L., Zhang, L., Wang, X., Diankov, G. & Dai, H. *Nature* **458**, 877–880 (2009).
9. Campos-Delgado, J. *et al.* *Nano Lett.* **8**, 2773–2778 (2008).
10. Cano-Márquez, A. G. *et al.* *Nano Lett.* **9**, 1527–1533 (2009).
11. Ci, L. J. *et al.* *Nano Res.* **1**, 116–122 (2008).
12. Meneses-Rodríguez, D. *et al.* (personal communication).

Tbx18 and the fate of epicardial progenitors

Arising from: C.-L. Cai *et al.* *Nature* **454**, 104–108 (2008)

Uncovering the origins of myocardial cells is important for understanding and treating heart diseases^{1,2}. Cai *et al.*³ suggest that *Tbx18*-expressing epicardium provides a substantial contribution to myocytes in the ventricular septum and the atrial and ventricular walls. Here we show that the T-box transcription factor gene 18 (*Tbx18*) itself is expressed in the myocardium, showing that their genetic lineage tracing system does not allow conclusions of an epicardial origin of cardiomyocytes *in vivo* to be drawn.

During amniote embryogenesis, cells from the proepicardium, a mesothelial cell cluster located at the venous pole of the heart, emigrate onto the myocardium to form the epicardium—the epithelial outer lining of the heart. Lineage-tracing studies in chick demonstrated that epicardial cells invade the myocardium, giving rise to coronary vascular support cells and adventitial fibroblasts^{4–6}.

In mouse, genetic lineage studies can be used to assess the contribution of precursor cells to a mature tissue or organ⁷. For the proepicardium/epicardium, such an effort has now been undertaken by Cai *et al.*³ based on the observation that the transcription factor gene *Tbx18* is expressed at high levels in these embryonic tissues⁸. Using a knock-in of the Cre recombinase gene in the *Tbx18* locus, epicardial cells were labelled irreversibly and their daughters were followed during development. The authors found that, in addition to known fates, *Tbx18*-positive epicardial cells contributed to the myocardium of the interventricular septum (IVS) and myocardial wall. Clearly, such a finding would define the epicardium as a previously unknown progenitor pool of cardiomyocytes complementing the earlier contributions of the first and second heart field⁹.

Prerequisite to any genetic lineage study is the specificity of the gene that drives Cre recombinase. In this specific case, *Tbx18* (*Cre*) should not be expressed in the myocardium. In previous studies, we observed robust *Tbx18* expression in the IVS and left ventricular wall at embryonic day (E)12.5 (refs 10, 11). Further analysis revealed continuous *Tbx18* expression in the IVS and the myocardium of the left ventricle from E10.5 to at least E16.5 (Fig. 1a–d, f, g). *Tbx18* expression co-localized with myocardial (*Tnni3*) staining on adjacent sections (Fig. 1e). Immunofluorescent analysis using an anti-*Tbx18* antibody confirmed that *Tbx18* protein is continuously expressed in the IVS and the left ventricular cardiomyocytes (Nkx2-5-positive) from E10.5 onwards (Fig. 1i–m). Detection of pre-messenger RNA in epicardium-free IVS myocardium tissue obtained by laser capture microdissection from E11.5 wild-type hearts confirmed that *Tbx18* expression is initiated *de novo* in cardiomyocytes (Fig. 1h). A protein kinase C iota (*Prkci*) null mutant (a gift from M. Leitges), which completely fails to form epicardium around the ventricles and dies at E12.5, expressed *Tbx18* in the IVS in a pattern similar to that in control embryos (Fig. 1n–q), further demonstrating that *Tbx18* expression in the myocardium of the IVS and left ventricle is independent from that in the epicardium.

We have generated a Cre knock-in allele of *Tbx18* (*Tbx18*^{Cre}), the Cre activity pattern of which faithfully reflects endogenous *Tbx18* expression⁸ (Fig. 2a). Comparison of expression of *Tbx18* from the wild-type allele and *lacZ* from the R26K^{lacZ} reporter¹² on adjacent sections of hearts of *Tbx18*^{Cre/+};R26K^{lacZ} embryos showed overlap of the Cre recombination pattern (*lacZ* expression) and endogenous *Tbx18* in the epicardium and in the IVS and left ventricular wall (Fig. 2b, c). Furthermore, we observed unchanged Cre activity from the *Tbx18*^{Cre} allele in the left ventricle and the septum region in mice mutant for the Wilms tumor 1 (*Wt1*) gene¹³ (Fig. 2d, e). Formation of epicardium and subepicardial mesenchyme is severely compromised in these mice¹⁴, again indicating that myocardial *Tbx18* expression is independent from the presence and functionality of the epicardium.

Together, our results unambiguously show that *Tbx18* expression is also present in cardiomyocytes of the IVS and the ventricular wall from E10.5 onwards.

The failure of Cai *et al.* to detect expression of *Tbx18* in the myocardium may result from lack of sensitivity in their expression assays and tools (nlacZ/nGFP knock-in allele)³. Their *Tbx18*^{Cre} allele is likely to be active in cardiomyocytes, similar to our *Tbx18*^{Cre} allele, but its pattern may be affected by the presence of the PGK-neo cassette³ (compare dissimilarity in recombination in somites and head region in Fig. 2b in ref. 3 and Fig. 2a here), which could lead to ectopic Cre activity in the heart. A genetic lineage study similar to that in ref. 3, following the fate of *Wt1*^{Cre}-expressing epicardium, also indicated that epicardium differentiates to myocardium¹⁵. Our analysis,

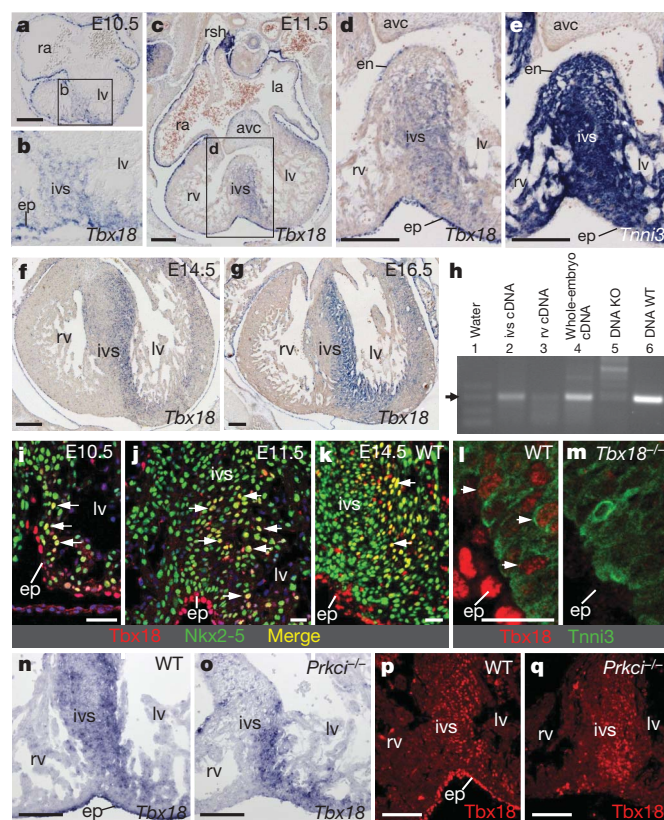


Figure 1 | Expression of *Tbx18* is not confined to the epicardium but is also found in cardiomyocytes of the interventricular septum and the left ventricle. a–g, Section *in situ* hybridization analysis for *Tbx18* (a–d, f, g) and cardiac troponin I (*Tnni3*, e) expression on paraffin sections of E10.5 (a, b), E11.5 (c–e), E14.5 (f) and E16.5 (g) hearts. Insets (a, c) are magnified (b, d). h, Pre-mRNA of *Tbx18* is detected in laser-captured epicardium-free IVS from E11.5 wild-type (WT) hearts by RT–PCR analysis for intron 1 sequences (lane 2). Background signal is observed in tissue from the right ventricle of the same hearts (lane 3). KO, knockout. i–m, Immunofluorescent analysis of *Tbx18* protein expression on paraffin sections of E10.5 (i), E11.5 (j, l, m) and E14.5 (k) hearts. Coexpression analyses of Nkx2-5 (yellow nuclei, white arrows in i–k). Absence of *Tbx18* protein signals in sections of *Tbx18*^{GFP/GFP} hearts (m). n–q, Section *in situ* hybridization analysis of *Tbx18* mRNA (n, o) and immunofluorescent analysis of *Tbx18* protein (p, q) in E12.5 *Prkci* mutant hearts that lack an epicardium (o, q). avc, atrioventricular cushion tissue; en, endocardium; ep, epicardium; ivs, interventricular septum; la, left atrium; lv, left ventricle; ra, right atrium; rsh, right sinus horn; rv, right ventricle. Scale bars: 100 μm (a–g, n–q); 25 μm (i–m).

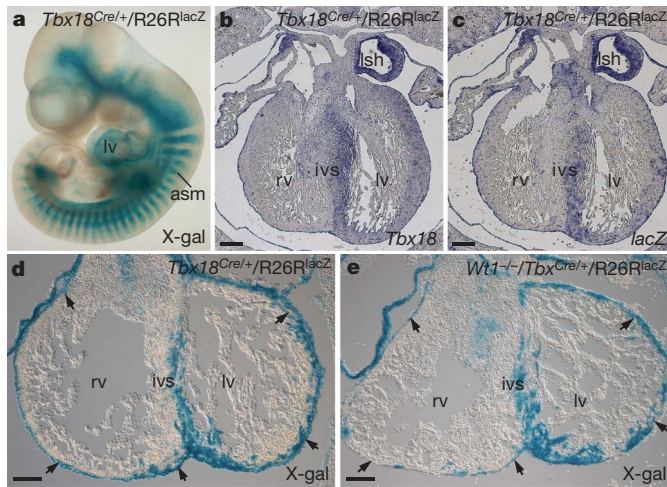


Figure 2 | Cre activity from a Cre-knock-in allele of *Tbx18* (*Tbx18^{Cre}*) recapitulates endogenous *Tbx18* expression. *Tbx18^{Cre}* mice were crossed to *R26R^{lacZ}* indicator mice¹² and expression and activity of the reporter was detected in compound *Tbx18^{Cre/+}/R26R^{lacZ}* embryos. **a**, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) staining on a whole-mount E10.5 embryo faithfully reflects endogenous *Tbx18* expression⁸. **b, c**, Section *in situ* hybridization analysis on adjacent paraffin sections of E14.5 hearts shows complete overlap of *Tbx18* (**b**) and *lacZ* reporter gene (**c**) expression. **d, e**, X-gal staining for *Tbx18*-driven Cre activity on transverse cryo-sections from E12.5 hearts of wild-type embryos (**d**) and a homozygous mutant for *Wt1* (**e**, ref. 14). The formation of the epicardium (arrows) is severely compromised in mutants. asm, anterior somite halves; lsh, left sinus horn. Other abbreviations are as in Fig. 1. Scale bars, 100 μm (**b–e**).

which shows that the genetic lineage study by Cai *et al.*³ does not unequivocally demonstrate an epicardial origin of cardiomyocytes *in vivo*, supports the notion that independent approaches are required to assess the relationship of a specific progenitor cell to a specific differentiated cell type.

Vincent M. Christoffels¹, Thomas Grieskamp², Julia Norden², Mathilda T. M. Mommersteeg¹, Carsten Rudat² & Andreas Kispert²

¹Department of Anatomy & Embryology, Heart Failure Research Center, Academic Medical Center, University of Amsterdam, Meibergdreef 15 L2-108, 1105 AZ Amsterdam, The Netherlands.

²Institut für Molekularbiologie, OE5250, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, 30625 Hannover, Germany.

e-mail: kispert.andreas@mh-hannover.de

Received 1 July 2008; accepted 3 February 2009.

- Olson, E. N. A decade of discoveries in cardiac biology. *Nature Med.* **10**, 467–474 (2004).
- Murry, C. E., Field, L. J. & Menasche, P. Cell-based cardiac repair: reflections at the 10-year point. *Circulation* **112**, 3174–3183 (2005).
- Cai, C. L. *et al.* A myocardial lineage derives from *Tbx18* epicardial cells. *Nature* **454**, 104–108 (2008).
- Mikawa, T. & Fischman, D. A. Retroviral analysis of cardiac morphogenesis: discontinuous formation of coronary vessels. *Proc. Natl Acad. Sci. USA* **89**, 9504–9508 (1992).
- Mikawa, T. & Gourdie, R. G. Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with ingrowth of the epicardial organ. *Dev. Biol.* **174**, 221–232 (1996).
- Winter, E. M. & Gittenberger-de Groot, A. C. Epicardium-derived cells in cardiogenesis and cardiac regeneration. *Cell. Mol. Life Sci.* **64**, 692–703 (2007).
- Yamauchi, Y. *et al.* A novel transgenic technique that allows specific marking of the neural crest cell lineage in mice. *Dev. Biol.* **212**, 191–203 (1999).
- Kraus, F., Haenig, B. & Kispert, A. Cloning and expression analysis of the mouse T-box gene *Tbx18*. *Mech. Dev.* **100**, 83–86 (2001).
- Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. *Nature Rev. Genet.* **6**, 826–835 (2005).
- Franco, D. *et al.* Left and right ventricular contributions to the formation of the interventricular septum in the mouse heart. *Dev. Biol.* **294**, 366–375 (2006).
- Bakker, M. L. *et al.* Transcription factor *Tbx3* is required for the specification of the atrioventricular conduction system. *Circ. Res.* **102**, 1340–1349 (2008).
- Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
- Kreidberg, J. A., Sariola, H., Loring, J. M., Maeda, M. & Pelletier, J. Housman, D. Jaenisch, R. *WT-1* is required for early kidney development. *Cell* **74**, 679–691 (1993).
- Moore, A. W., McInnes, L., Kreidberg, J., Hastie, N. D. & Schedl, A. YAC complementation shows a requirement for *Wt1* in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* **126**, 1845–1857 (1999).
- Zhou, B. *et al.* Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* **454**, 109–113 (2008).

doi:10.1038/nature07916

Cai *et al.* reply

Replying to: V. M. Christoffels *et al.* *Nature* **458**, doi:10.1038/nature07916 (2009)

We recognized the importance of confirming data from Cre lineage studies with an alternative approach, and thus used a fluorescent dye to lineage-label epicardial cells. This approach confirmed that epicardial cells contributed to myocytes within the heart¹. In our experiments, we never observed *Tbx18* mRNA within ventricles at embryonic day E10.5 to E11.5, although our RNA signal in epicardium at these stages is at least as strong as that observed by Christoffels *et al.*². We note a discrepancy between amount of RNA staining (broadly throughout the middle of the ventricle) and protein staining (much more restricted) that Christoffels *et al.*² observe at early stages. Reasons for discrepancies between our RNA *in situ* data and that of Christoffels *et al.* are unclear. Christoffels *et al.*² detect *Tbx18* intronic sequences by RT-PCR of E11.5 IVS, but do not provide data to demonstrate mature *Tbx18* mRNA. Notably, we did observe active expression of *Tbx18* in non-myocytes after E11.5 (ref. 1).

Expression of *Tbx18:nlacZ* in our experiments is consistent with our RNA *in situ* data for *Tbx18*. Despite overstaining, and examination of multiple hearts and sections, *Tbx18:nlacZ* was not expressed within the heart at E10.5 and E11.5—times at which we observed lineage-traced *Tbx18:Cre* myocardial cells¹. Immunostaining for β-galactosidase (β-gal) protein was consistent with X-gal data.

Why there are differences between these data and the Christoffels *et al.*² *Tbx18* immunostaining data is unclear. *Tbx18* antibody staining in heart by Christoffels *et al.*² demonstrates much lower levels of positive staining for *Tbx18* within heart (barely above background) than in epicardium (in which it is robust). Data presented on *Prkci*^{−/−} mutants are difficult to comment on, as this model has not been described in the literature. Published *Prkci*^{−/−} mutants do not survive after E9.5 (ref. 3). At E12.5, we also observe active expression of *Tbx18:nlacZ* within the heart, but have found that actively expressing cells are non-myocytes. Christoffels *et al.* do not address the cell type observed in *Prkci* mutants.

With regard to *Tbx18* lineage studies in the *Wt1* mutant background, it is important to consider that the *Wt1* mutant phenotype is variable and dependent on strain background⁴. Mutants can die as early as E12.5 or survive throughout gestation. Even on an inbred C57/BL6 background, lethality occurs over a relatively broad window from E12.5 to E15 (refs 4–6). Even in the most severely affected cases, the epicardium forms throughout much of the heart (see Fig. 2b of ref. 6), including the caudal region adjacent to the septum and the left ventricle, and entry of epicardial cells into the heart is observed in regions of interventricular and atrioventricular sulci⁶. In *Wt1* mutant

data presented by Christoffels *et al.*², there are a substantial number of epicardial cells present, notably adjacent to most lineage-traced cells within the interventricular region. Segments lacking epicardial cells could reflect initial absence in these regions, or poor attachment or maintenance of epicardial cells that may previously have undergone inward migration. From the foregoing observations, it is unknown how much migration has occurred in these *Wt1* mutants, making it difficult to draw conclusions from this study.

We feel that overall the evidence is consistent with a contribution of *Tbx18* epicardial cells to myocardial cells. Lineage-tracing of epicardium by fluorescent dye and inducible *Wt1-Cre* by Zhou *et al.* have also demonstrated a contribution of epicardial cells to myocytes⁷.

Chen-Leng Cai^{1,4,5,6}, **Jody C. Martin**², **Yunfu Sun**¹, **Li Cui**¹,
Lianchun Wang⁷, **Kunfu Ouyang**³, **Lei Yang**¹, **Lei Bu**¹, **Xingqun Liang**³,
Xiaoxue Zhang¹, **William B. Stallcup**⁸, **Christopher P. Denton**⁹,
Andrew McCulloch², **Ju Chen**³ & **Sylvia M. Evans**^{1,3}

¹Skaggs School of Pharmacy, University of California, San Diego, La Jolla, California 92093, USA.

e-mail: syevans@ucsd.edu

²Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA.

³Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA.

⁴Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029, USA.

⁵Centre for Molecular Cardiology, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029, USA.

⁶The Black Family Stem Cell Institute, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029, USA.

⁷Complex Carbohydrate Research Centre, University of Georgia, Athens, Georgia 30602, USA.

⁸Developmental Neurobiology Programme, The Burnham Institute, La Jolla, California 92037, USA.

⁹Centre for Rheumatology, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, UK.

1. Cai, C. L. *et al.* A myocardial lineage derives from *Tbx18* epicardial cells. *Nature* **454**, 104–108 (2008).
2. Christoffels, V. M. *et al.* *Tbx18* and the fate of epicardial progenitors. *Nature* **458**, doi:10.1038/nature07916 (2009).
3. Soloff, R. S., Katayama, C., Lin, M. Y., Feramisco, J. R. & Hedrick, S. M. Targeted deletion of protein kinase C lambda reveals a distribution of functions between the two atypical protein kinase C isoforms. *J. Immunol.* **173**, 3250–3260 (2004).
4. Herzer, U., Crocoll, A., Barton, D., Howells, N. & Englert, C. The Wilms tumor suppressor gene *Wt1* is required for development of the spleen. *Curr. Biol.* **9**, 837–840 (1999).
5. Kreidberg, J. A. *et al.* *WT-1* is required for early kidney development. *Cell* **74**, 679–691 (1993).
6. Moore, A. W., McInnes, L., Kreidberg, J., Hastie, N. D. & Schedl, A. YAC complementation shows a requirement for *Wt1* in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* **126**, 1845–1857 (1999).
7. Zhou, B. *et al.* Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* **454**, 109–113 (2008).

doi:10.1038/nature07917

The dawn of the particle astronomy era in ultra-high-energy cosmic rays

Pablo M. Bauleo^{1*} & Julio Rodríguez Martino^{2,3*}

Cosmic rays are charged particles arriving at the Earth from space. Those at the highest energies are particularly interesting because the physical processes that could create or accelerate them are at the limit of our present knowledge. They also open the window to particle astronomy, as the magnetic fields along their paths are not strong enough to deflect their trajectories much from a straight line. The Pierre Auger Observatory is the largest cosmic-ray detector on Earth, and as such is beginning to resolve past observational disagreements regarding the origin and propagation of these particles.

In 1912, after a series of balloon flights, Hess discovered a penetrating radiation that originated in outer space. Years later, in 1926, Millikan called this radiation ‘cosmic rays’. The name has survived since then, generally referring to charged particles impinging on the Earth’s atmosphere. In the late 1930s, Auger and his group measured coincident signals generated by detectors separated by distances of more than a few hundred metres^{1,2}; they concluded that these signals were caused by an ‘extensive air-shower’ (EAS) of charged particles. Auger and his co-workers assumed that the air-shower was originated by a single photon, high in the atmosphere, and used the recently developed quantum electrodynamics theory to estimate its energy, which they found to be in excess of 10^{15} electron volts (eV). Figure 1 shows a schematic representation of an EAS.

Cosmic rays of energies larger than about 10^{13} eV are small in number, and so can only be detected through the secondary particles produced when they enter Earth’s atmosphere. The EAS starts with the interaction of a cosmic ray with a nucleus in the upper atmosphere. All the available energy is distributed among the secondary particles—of which there can be billions if the primary energy is above 5×10^{17} eV—that can spread over several tens of square kilometres at ground level. Two methods are mainly used to register these particle cascades. The particle density can be sampled at the ground using an array of detectors; alternatively, the shower path can be tracked through the atmosphere, collecting the fluorescence light induced by electrons in the atmospheric nitrogen molecules.

Here we review the developments in ultra-high-energy cosmic ray (UHECR) physics over the past 15 years: we cover the controversy about the existence of the theoretically predicted suppression of the cosmic-ray energy spectrum, and its later confirmation. The most relevant topic is the discovery that the arrival direction of the most energetic cosmic rays follows the distribution of nearby extragalactic objects. This implies that their origin is not cosmological, but instead they are accelerated inside extragalactic objects, by some still unclear physical process. Three large experimental facilities—AGASA (Akeno Giant Air Shower Array), HiRes (High Resolution Fly’s Eye) and the Pierre Auger Observatory—have already started what will eventually become a new era in astronomy. In the near future, further observations and more accurate instruments will identify the cosmic-ray acceleration sites and will lead to the study of the energy spectrum of individual sources. This, combined with the study of the attenuation of cosmic rays through space, could give valuable

information on the cosmic microwave background. The deflection produced on the cosmic-ray path by Galactic and extragalactic magnetic fields will be an indirect tool to measure their strength. In addition, accurate measurements of the interaction of cosmic rays with the Earth’s atmosphere will hint at the particle physics interaction models, at an energy range beyond what can be achieved in

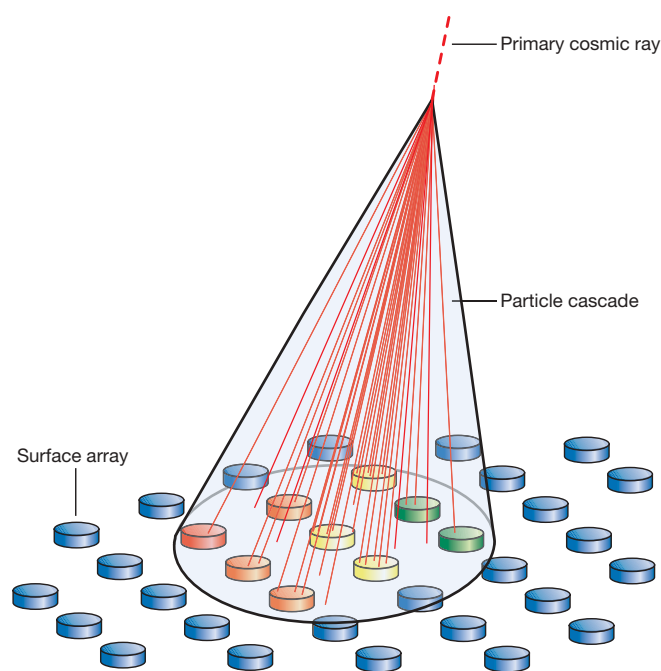


Figure 1 | Scheme of an extensive air-shower. The primary cosmic ray (dashed line) undergoes a nuclear interaction in the upper atmosphere (typically 20 km above sea level), producing a cascade of elementary particles (represented as solid red lines within a conical shape). These particles propagate across the atmosphere and could reach ground level. The cascade footprint at the ground could be of tens of square kilometres. A network of particle detectors at ground level (surface array) can detect the arrival of the particles, allowing reconstruction of the whole cascade. Different colours in the scheme represent different arrival times of the particles.

¹Physics Department, Colorado State University, Fort Collins, Colorado 80523, USA. ²Dipartimento di Fisica e Astronomia, Università di Catania, ³INFN Sezione di Catania, I-95123 Catania, Italy.

*These authors contributed equally to this work.

human-made accelerators. These observations are within reach of the current and next-generation observatories, and will herald the dawn of the era of charged-particle astronomy.

Properties of cosmic rays

The observed cosmic-ray energy spectrum spans from 10^8 eV to more than 10^{20} eV. Particles with energies lower than 10^{10} eV mainly come from the Sun, as the solar wind prevents particles in that energy range from reaching the Earth from outside the Solar System. For energies higher than 10^{18} eV, a convincing explanation of the acceleration processes and sources is still unknown. Some theories suggest that these cosmic rays originate in stellar winds within our Galaxy and later accelerate in supernova shocks or similar high-energy environments³. Active galactic nuclei (AGN), galaxies with very intense emission in a broad wavelength range, are possible source candidates of UHECRs above 10^{19} eV (ref. 4), but so far there are only experimental hints suggesting this.

The cosmic-ray flux follows a power law ($E^{-\gamma}$) as a function of energy E , with an approximate index $\gamma = 3$. This index value remains remarkably constant, showing only small variations across the whole measured cosmic-ray energy spectrum.

At the highest energies, above 10^{20} eV, the estimated number of particles is only a few per km^2 per millennium. This extremely low flux calls for the construction of huge observatories, covering a very large area with detectors. For instance, the Pierre Auger Southern Observatory in Argentina covers $3,000 \text{ km}^2$, which is about 30 times the size of the district of Paris.

Cosmic rays with energies above 4×10^{19} eV cannot travel through space without being attenuated^{5,6}. Propagation is mainly affected by the presence of the cosmic microwave background radiation, consisting of photons with a black-body radiation distribution corresponding to an equivalent temperature of 2.7 K. In the rest frame of an extremely energetic proton, these low-energy photons are seen as very high energy photons (γ -rays), of about 10^8 eV. If the photon energy in the rest frame is above 150 MeV, pion-production reactions become possible. The proton loses energy in each reaction, reducing the mean distance it can travel undisturbed to about 50 Mpc. This effect produces a dip in the spectrum, known as the 'GZK suppression' (named after Greisen, Zatsepin and Kuzmin, who predicted its existence), and it reduces the number of high energy particles able to arrive at Earth, if originated at larger distances.

The HiRes observatory data suggested the presence of suppression in the flux of cosmic rays in the GZK energy region⁷, whereas the AGASA collaboration announced that the cosmic-ray spectrum continued, with a power law dependence, above GZK energies⁸. This last result was revisited a few years later, without being able to arrive at a definite conclusion owing to the limited number of events in the GZK energy region, even though the existing, limited data collected by AGASA is still being re-analysed^{9,10}.

The Pierre Auger Southern Observatory data seem to agree with the HiRes result in the GZK energy region¹⁰, resolving the controversy between the two previous experimental results. The observation of the GZK suppression¹¹ is another interesting result. The larger data set of the Auger Observatory made it also possible to establish a correlation between some high energy events and AGN (or any other astronomical objects that follow the same spatial distribution) closer than 75 Mpc to the Earth¹². For protons with energy larger than 6×10^{19} eV, the magnetic deflection of the trajectory of the cosmic rays is only a few degrees¹², hence enabling the possibility of particle astronomy. This small deflection would imply that the particles 'point back' to their sources, making it possible to identify the origin of cosmic rays and even study the spectra of individual sources. By studying the distribution of cosmic-ray arrival directions (such as clustering, thread-like structures, and so on), it would be possible to analyse the properties of Galactic and inter-galactic magnetic fields.

The AGASA observatory

AGASA¹³ was located at Akeno, Japan. It ran in full operation mode from 1993 to 2004, being able to take data continuously, independently of weather conditions. Each ground station, composed of plastic scintillators, sampled the secondary particles of an EAS. The trigger time of each individual station was used in the reconstruction of the EAS arrival direction, while the energy measurement was based on the number of particles at each station.

The energy of a shower detected using a ground array is not measured directly. The particle density at a given distance from the EAS axis is correlated with the energy of the primary cosmic ray through computer simulations. The models used in the simulations are based on the knowledge about interactions acquired in particle accelerator experiments. This means that the models extrapolate the physical processes to several orders of magnitude in energy beyond what has been measured until now. One of the Large Hadron Collider experiments (LHCf) will be dedicated to reducing the uncertainty in hadron interaction models of cosmic-ray showers, by measuring the forward particle production in proton interactions¹⁴. Apart from this, the computational effort of producing and tracking about 10^{11} particles is too large to be practical. Hence, only a statistically representative sample of the EAS secondary particles is propagated to the ground in the simulation. All these facts lead to an energy measurement that is strongly model dependent, and to large uncertainties in its value.

The HiRes observatory

The HiRes Fly's Eye¹⁵ was located in Dugway, Utah, USA. HiRes commissioned its first location in 1997 and its second location in 1999. Both locations were decommissioned in 2006. This observatory collected fluorescence light induced in the atmosphere by the passing EAS. The total brightness of an EAS, in fluorescence light, averages a few watts. The amount of light collected is so faint that these detectors can only operate on clear, moonless nights. Typical observation duty factors of fluorescence detectors lie between 10% and 15%.

Each of the HiRes locations had mirrors that focused the fluorescence photons into a light sensor array, or 'camera'. It is conceptually similar to a CCD camera, with each light sensor playing the role of one pixel. By considering the relative trigger times and geometric pattern of the pixels in the camera, it is possible to reconstruct the arrival direction of the shower. The energy is calculated by integrating the total amount of light measured at the detector location. The total number of photons induced in the atmosphere by the EAS is proportional to the total available energy, that is, the energy of the primary cosmic ray. Some particles in the cascade do not induce fluorescence light and the total energy of the EAS must be corrected to account for this fact. The atmospheric conditions are other factors to include when estimating the primary cosmic-ray energy. An atmospheric attenuation correction, based on the distance from the EAS to the detector, needs to be applied.

Discrepant results

The limited sample of cosmic rays in the GZK energy region, together with intrinsic differences in the way each experiment measured the cosmic-ray energy, set the stage for a controversial difference between the measured spectra.

A comparison of both measured spectra is shown in Fig. 2, where the discrepancy is clear. The AGASA data seemed to favour the absence of a suppression, while HiRes spectrum followed the expected curve. Both results should be interpreted carefully, as the calculations involved are not straightforward and, again, the number of detected events was not enough to firmly establish either claim¹⁶.

When computing a cosmic-ray spectrum, it is critical to calculate the instrument exposure, or time-integrated collection area. In the case of AGASA, the exposure is reduced to the convolution of the detector array geometrical area, the acceptance solid angle and the effective running time. The acceptance of a surface array, like AGASA, becomes

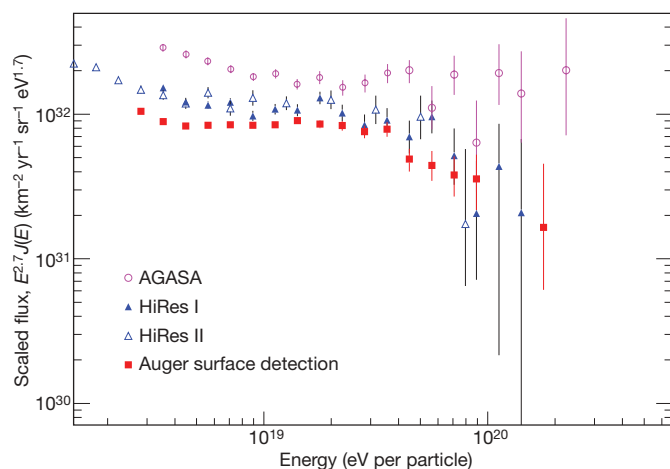


Figure 2 | UHECR data from different experiments. Comparison between AGASA (circles), HiRes monocular spectra (open and filled triangles correspond to each HiRes location) and Pierre Auger Southern Observatory. Error bars are 1σ .

constant with energy, once its trigger efficiency reaches 100%. The cumulative exposure for this detector is about $1,600 \text{ km}^2 \text{ sr yr}$ (ref. 17).

A fluorescence detector requires a more complicated exposure calculation. The collection volume is a hemisphere (centred at the detector location), the radius of which indicates the maximum observation distance for a given EAS. This distance changes with the atmospheric conditions (atmospheric aerosols, cloud coverage, position of the clouds) and depends on the EAS energy. The acceptance of a fluorescence detector, like HiRes, is a function of the EAS energy (the brighter the EAS, the further away it can be detected). This implies that, in order to calculate the energy spectrum of an EAS, it is necessary to accurately know how the instrument acceptance depends on the EAS energy and the atmospheric conditions at the time of measurement. It is very difficult to deduce the exposure of HiRes from the published results, but it is quoted as “more than twice that of AGASA above the GZK-threshold”⁷⁷.

In any case, both these experiments have statistically limited data samples, given the extremely low cosmic-ray flux at those energies¹⁶. In response to the AGASA results, numerous speculations about how cosmic rays could avoid energy loss on their way to Earth were proposed. Either new particles^{18,19} or interactions with magnetic fields²⁰ were invoked to avoid the problem. These articles are just a small sample of a long list showing different (and sometimes quite ingenious) arguments.

Besides measuring the energy spectrum, both experiments analysed the arrival direction distributions of cosmic rays. An ‘ n -plet’ is defined as a set of n independent events whose arrival directions are the same, within experimental uncertainties. The AGASA collaboration found 5 doublets and 1 triplet²¹ in their data sample, where only 2 doublets were expected statistically. This result was not confirmed by HiRes²². On the other hand, correlations were found in the HiRes sample with the locations of BL Lacertae objects (AGN with their jets pointing towards Earth)²³, although they have not been confirmed by an independent data sample. It should be remarked that anisotropy in the arrival direction of cosmic rays is not expected at lower energies. However, at higher energies—combining data from different observatories—an excess of events coming from the supergalactic plane (a plane defined by the locations of the galaxies in the local cluster) was found for events with energies above $4 \times 10^{19} \text{ eV}$, giving a hint that the origin of UHECRs is most likely to be extragalactic²⁴. This result was independently suggested later by analysing the shape of the cosmic-ray spectrum^{25,26}.

The relatively low exposures of these experiments could only provide hints about the arrival direction of the cosmic rays, making it possible to search for clustering and sources, but not to confirm

them. Still, these results were extremely important in that they showed anisotropy studies (and potentially the identification of cosmic-ray sources) to be within reach.

The Pierre Auger Observatory

The Pierre Auger Southern Observatory²⁷, schematically shown in Fig. 3, is located in the province of Mendoza, Argentina. It covers an approximate area of $3,000 \text{ km}^2$, which makes it the largest cosmic-ray observatory to date. Its northern counterpart will be built in the vicinity of Lamar, Colorado, USA. When finished, the joint instruments will have full sky coverage as observed from both hemispheres. The Southern Observatory has been collecting cosmic-ray data since 2004, while increasing its size up to the installation of the last surface detector on June 2008. As of 31 August 2007, the accumulated exposure of the Southern Observatory is $9,000 \text{ km}^2 \text{ sr yr}$ (ref. 28). The yearly accumulated exposure is about $6,000 \text{ km}^2 \text{ sr yr}$ and the observatory is expected to operate for a total of 20 years.

This observatory combines the techniques used in previous experiments, by means of a ‘hybrid detector’, that is, having a fluorescence detector and an array of surface detectors working together. The fluorescence detector follows the shower cascade across the atmosphere and the surface detector array—in this case water Cherenkov detectors—detects the particles on arrival at ground level. Hybrid measurements can set an absolute energy scale, improve the determination of the primary particle type and give better energy and angular resolution²⁷. This approach provides a model-independent energy calculation, using the fluorescence detector data together with the simple surface array aperture calculation.

In hybrid mode, for any given EAS measured simultaneously by both instruments, the energy deposited in the atmosphere—as recorded by the fluorescence detector—is then related to a surface detector energy parameter. Then, this model-independent correlation can be used as energy calibration for events measured only with the surface detector array²⁹.

The Pierre Auger Collaboration is taking advantage of the unique characteristics of the observatory. Although the limit is arbitrary, EAS detected by water Cherenkov arrays are typically reconstructed only up to 60° . Auger Collaboration members have developed analysis techniques to extend the acceptance up to 75° , which increases the

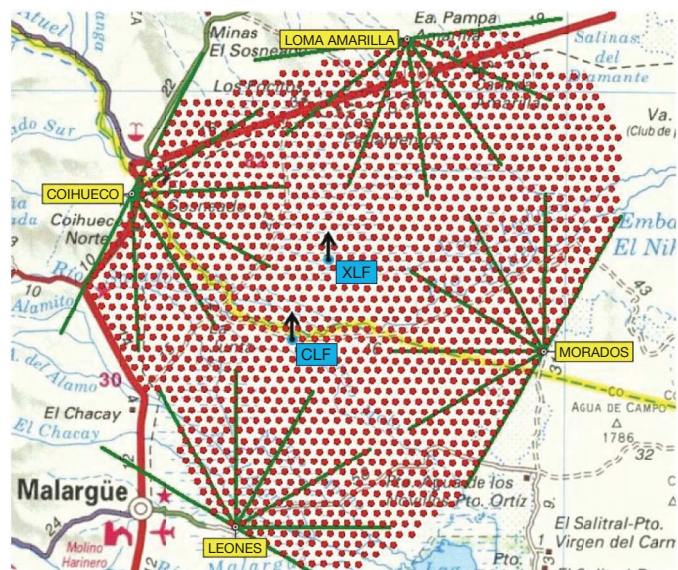


Figure 3 | The Pierre Auger Southern Observatory. It consists of an array of 1,600 surface detectors (red filled circles), complemented by 24 fluorescence detectors, grouped in four buildings (yellow labels; Leones, Morados, Loma Amarilla and Coihueco). Green lines represent the field of view of each detector. Two laser facilities (blue labels; CLF and XLF) are available for energy calibration and atmospheric monitoring. Observatory information is superimposed on a map of the area.

instrument exposure even further³⁰. As a comparison, plastic scintillator arrays (like AGASA, or the Telescope Array in Utah, see below) typically limit their reconstructed events up to 45°.

Evidence for the GZK suppression

Results recently published by the Auger Collaboration¹¹ report the existence of a deficit of cosmic rays at the highest energies. Still, this result alone is not enough as to prove that the GZK suppression has been observed. It could be that the energy spectrum is limited by the maximum energy available at the cosmic-ray acceleration sites.

When the evidence on the deficit in the flux of cosmic rays is put together with the energy at which the correlation with nearby extragalactic objects²⁸ sets in, one could then argue that the GZK suppression has been observed. If objects beyond an approximate distance of 75 Mpc were to be included in the analysis, the correlation would very rapidly diminish.

Although both HiRes and the Pierre Auger Southern Observatory have observed a suppression in the cosmic-ray flux above an energy of approximately 4×10^{19} eV, differences still exist in the measured spectrum index and the overall energy normalization. The energy scales of these two observatories differ by about 17% (ref. 31).

The sources

One of the main questions to be answered regarding UHECRs is how these particles can be accelerated to such energies. The size of the acceleration region and the magnetic field present in it must follow a relation, usually represented in a Hillas plot like that shown in Fig. 4. Only a few astrophysical objects could then be potential sources.

Arguably, the most relevant recent observation has been the discovery of a correlation between cosmic-ray arrival directions and nearby extragalactic objects^{12,28}. The correlation found in the Pierre Auger Southern Observatory data becomes significant for cosmic rays above 5.7×10^{19} eV and AGN within 75 Mpc. With those

parameters, 20 events (out of a total of 27) lie within 3.1° from an object listed in the Veron-Cetty-Veron catalogue³².

AGN have traditionally been considered as possible candidates for cosmic-ray acceleration sites. However, any other astrophysical object close enough to Earth to avoid the GZK suppression, with a spatial distribution similar enough to that of AGN, could be the source.

The AGN hypothesis seems to be supported by the correlation found between the arrival direction of cosmic rays reported by the Auger Collaboration¹² and the positions of the Swift hard X-ray catalogue of AGN, when weighted by the X-ray flux and constrained to distances less than 100 Mpc (ref. 33). At the same time, using the same events measured by the Pierre Auger Southern Observatory, a correlation was also found with the HIPASS catalogue of H I spiral galaxies (when weighted by their H I flux)³⁴. The latter results do not contradict the correlation found with AGN, as all these objects trace the distribution of matter. The hypothesis of H I galaxies as cosmic-ray sources is interesting, as it would explain the lack of events from the Virgo cluster (which is not rich in H I galaxies).

HiRes members have searched their data for correlations³⁵ based on the Pierre Auger Southern Observatory parameters, and their analysis does not support the result published by the Auger Collaboration. Reference 31 shows that if corrected by the energy mismatch between both experiments, HiRes would have only 5 events in their stereo data sample, which might not be enough as to establish or reject any correlation.

Open questions

Despite having measured a suppression in the spectrum compatible with the GZK suppression and arrival direction anisotropies (or perhaps because of those facts), some exciting and intriguing questions still remain to be solved.

Sources and acceleration models. Nearby extragalactic objects have been found to correlate with the arrival direction of cosmic rays, but it is not yet possible to study the energy spectrum of individual sources. Such a spectrum would lead to a better understanding of acceleration processes at the sources. At the same time, the search for other potential sources should continue. Cosmic rays could be generated by different astrophysical objects.

Energy spectrum. The GZK suppression is produced by the interaction of nucleons with photons, at energies higher than 4×10^{19} eV. At energies higher than 3×10^{20} eV, the interactions become much less probable. Hence, cosmic rays with those energies could propagate almost undisturbed through space, allowing the study of the Universe at extreme energies. This feature, predicted by quantum physics, is known as the 'GZK recovery'. Observing it would prove quantum physics at an energy range that has not been explored before. The lack of a GZK recovery could imply new physics.

Mass composition and particle physics. A very important point to be studied is the mass composition of cosmic rays. This will either prove or reject different acceleration and propagation models, which favour either light or heavy primary particles. Moreover, at these high energies, cosmic-ray interactions with atoms in the upper atmosphere are in the range of a few hundred TeV (in the centre of mass frame). Studies of shower development in the atmosphere (known as elongation rate) will give an opportunity to unveil features of hadronic interactions at these energies, which are more than one order of magnitude higher than those achievable by the Large Hadron Collider, the most powerful human-made particle accelerator³⁶.

Magnetic fields. Magnetic fields could be studied by looking at the arrival direction pattern of cosmic rays as a function of energy. If 'strings' of events were identified, their relative deviation at different energies would allow us to set limits (or possibly even measure) the strength of Galactic and extragalactic magnetic fields.

A larger set of events, measured with good resolution, will answer several questions. As it is true for so many scientific disciplines, the main problem to be solved regarding the study of UHECRs is obtaining a significantly larger number of events.

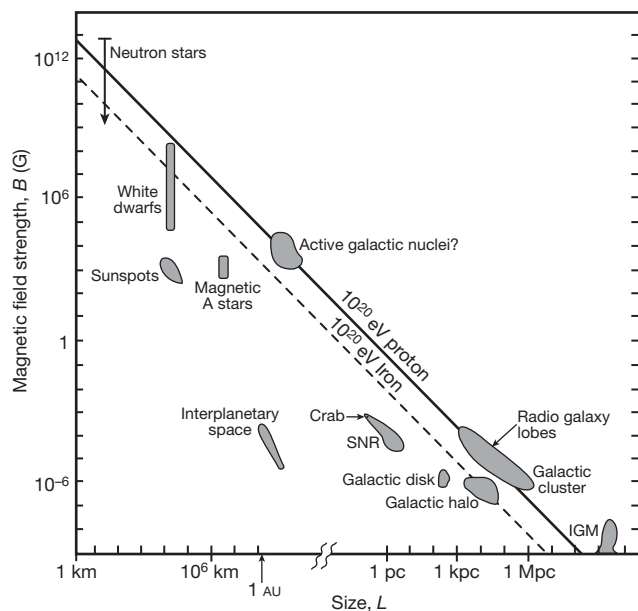


Figure 4 | Hillas diagram. Non-exotic acceleration processes require a particle to be confined within a region (of size L) where magnetic field shocks are present (with a field intensity value of B). Once the particle reaches its maximum energy, then the magnetic field is not able to keep the particle confined within the acceleration region and the particle escapes. This gives an approximate value for the maximum achievable energy of $E_{\text{max}} = BL$, shown as a solid/dashed line for a 10^{20} eV proton/iron nuclei, respectively. We show data for a variety of astrophysical objects; only those above the line can accelerate particles to energies into the GZK region. Crab indicates the Crab nebula; SNR, supernova remnant; IGM, intergalactic magnetic field.

The future

The Pierre Auger Southern Observatory is the largest-aperture observatory currently taking data and its exposure is larger than that of any previous detector. There is a proposal to increase the size of the projected Northern Observatory to cover an area 7 times larger than that of the Southern Observatory.

The Telescope Array in Utah, a hybrid instrument combining a surface scintillator array and fluorescence detectors, is the only observatory in the Northern Hemisphere currently taking data in this energy regime. Its yearly cumulative exposure will depend on the final operation conditions, but it could be estimated to be about $1,400 \text{ km}^2 \text{ sr yr}$ (ref. 37).

New techniques and observation methods are being considered. The collection of fluorescence light with space-based instruments, looking down into the Earth's atmosphere, has been proposed. JEM-EUSO³⁸ and OWL³⁹ are examples of this technique. Radio-wave detection of EAS is also currently being developed⁴⁰.

The past decade has proven fruitful and exciting in cosmic-ray physics. We have witnessed revisions and improvements in the instrumental techniques, which in turn have paid off by establishing the existence of the GZK suppression and by the discovery of anisotropies in the cosmic-ray arrival directions. In cosmic-ray physics, discoveries have been achieved by seeking the largest exposure possible. History has shown us that in this field, exposure matters.

In the near future, within 4 years or so, the Pierre Auger Southern Observatory should have observed about 100 events above $\sim 5 \times 10^{19} \text{ eV}$. In contrast, the proposed Pierre Auger Northern Observatory could be collecting the above-mentioned statistics every 9 months. In 20 years of combined operation, about 2,000 events (above $\sim 5 \times 10^{19} \text{ eV}$) could have been observed. Such data from the Northern and Southern Observatories could be used to accurately search for point sources, to study the energy spectra of different sources, and to understand Galactic and extragalactic magnetic fields, as well as to investigate and perhaps uncover particle physics beyond accelerator energies. A new window to the Universe has been opened; we are witnessing the dawn of the particle astronomy era.

1. Auger, P. & Maze, R. Extensive cosmic showers in the atmosphere. *C.R. Acad. Sci. II* **207**, 228–229 (1938).
2. Auger, P., Ehrenfest, P., Maze, R., Daudin, J. & Robley, A. F. Extensive cosmic-ray showers. *Rev. Mod. Phys.* **11**, 288–291 (1939).
3. Biermann, P. & Sigl, G. Introduction to cosmic rays. *Lect. Notes Phys.* **576**, 1–26 (2001).
4. Biermann, P., Gaisser, T. & Stanev, T. The origin of galactic cosmic rays. *Phys. Rev. D* **51**, 3450–3454 (1995).
5. Greisen, K. End to the cosmic ray spectrum? *Phys. Rev. Lett.* **16**, 748–750 (1966).
6. Zatsepin, G. T. & Kuzmin, V. A. Upper limit of the spectrum of cosmic rays. *Zh. Eksp. Teor. Fiz. Pisma Red.* **4**, 114–117 (1966).
7. Abbasi, R. et al. First observation of the Greisen-Zatsepin-Kuzmin suppression. *Phys. Rev. Lett.* **100**, 101101 (2008).
First observation of the cosmic-ray flux suppression at the highest energies.
8. Sakaki, N. et al. in *Proc. 27th Int. Cosmic Ray Conf. (ICRC 2001)* (eds Kampert, K. H., Hainzelmann, G. & Spierling, G.) 333–336 (Copernicus Gesellschaft, 2001).
9. Sokolsky, P. & Thomson, G. B. Highest energy cosmic rays and results from the HiRes experiment. Preprint at (<http://arXiv.org/abs/0706.1248v1>) (2007).
10. Watson, A. Recent results from the Pierre Auger Observatory – Including comparisons with data from AGASA and HiRes. *Nucl. Instrum. Meth. A* **588**, 221–226 (2008).
11. Abraham, J. et al. Observation of the suppression of the flux of cosmic rays above $4 \times 10^{19} \text{ eV}$. *Phys. Rev. Lett.* **101**, 061101 (2008).
Confirmation of the cosmic-ray flux suppression at the highest energies.

12. Abraham, J. et al. Correlation of the highest energy cosmic rays with nearby extragalactic objects. *Science* **318**, 938–943 (2007).
Discovery of cosmic-ray arrival direction anisotropy.
13. Chiba, N. et al. Akeno giant air shower array (AGASA) covering 100-km^2 area. *Nucl. Instrum. Meth. A* **311**, 338–349 (1992).
14. Pilkington, A. et al. in *Proc. XVI Int. Workshop on Deep-Inelastic Scattering and Related Topics (DIS 2008)* (eds Devenish, R. & Ferrando, J.) doi:10.3360/dis.2008.50 (Science Wise Publishing, 2008).
15. Abu-Zayyad, T. et al. The prototype high-resolution Fly's Eye cosmic ray detector. *Nucl. Instrum. Meth. A* **450**, 253–269 (2000).
16. De Marco, D., Blasi, P. & Olinto, A. V. On the statistical significance of the GZK feature in the spectrum of ultra high energy cosmic rays. *Astropart. Phys.* **20**, 53–65 (2003).
17. Takeda, M. et al. Energy determination in the Akeno giant air shower array experiment. *Astropart. Phys.* **19**, 447–462 (2003).
18. Kachelriess, M., Semikoz, D. V. & Tortola, M. A. New hadrons as ultra-high energy cosmic rays. *Phys. Rev. D* **68**, 043005 (2003).
19. Gibilisco, M. Monopolum decay as a source of ultrahigh energy cosmic rays. *Nucl. Phys. Proc.* **100** (Suppl.), 357–359 (2001).
20. Deligny, O., Letessier-Selvon, A. & Parizot, E. Magnetic horizons of UHECR sources and the GZK feature. *Astropart. Phys.* **21**, 609–615 (2004).
21. Teshima, M. et al. in *Proc. 28th Int. Cosmic Ray Conf. (ICRC 2003)* (eds Kajita, T., Asaoka, Y., Kawachi, A., Matsubara, Y. & Sasaki, M.) 437–440 (Universal Academy, 2003).
22. Bellido, J. et al. in *Proc. 28th Int. Cosmic Ray Conf. (ICRC 2003)* (eds Kajita, T., Asaoka, Y., Kawachi, A., Matsubara, Y. & Sasaki, M.) 425–428 (Universal Academy, 2003).
23. Abbasi, R. U. et al. Search for cross-correlations of ultra-high-energy cosmic rays with BL Lacertae objects. *Astrophys. J.* **636**, 680–684 (2006).
24. Uchiyori, Y. et al. Cluster analysis of extremely high energy cosmic rays in the northern sky. *Astropart. Phys.* **13**, 151–160 (2000).
25. Berezhinsky, V. On transition from galactic to extragalactic cosmic rays. *J. Phys. Conf. Ser.* **47**, 142–153 (2006).
26. Stanev, T. The transition from galactic to extragalactic cosmic rays. *Nucl. Phys. Proc.* **168** (Suppl.), 252–257 (2007).
27. Abraham, J. et al. Properties and performance of the prototype instrument for the Pierre Auger Observatory. *Nucl. Instrum. Meth. A* **523**, 50–95 (2004).
28. Abraham, J. et al. Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astropart. Phys.* **29**, 188–204 (2008).
29. Sommers, P. in *Proc. 29th Int. Cosmic Ray Conf. (ICRC 2005)* Vol. 7 (eds Sripathi Acharya, B. et al.) 387–390 (Tata Institute of Fundamental Research, 2005).
30. Newton, D. in *Proc. 30th Int. Cosmic Ray Conf. (ICRC 2007)* (eds Caballero, R. et al.) 323–325 (Universidad Nacional Autonoma de Mexico, 2007).
31. Kampert, K. H. Ultra high-energy cosmic ray observations. *J. Phys. Conf. Ser.* **120**, 062002 (2008).
32. Veron-Cetty, M. et al. A catalogue of quasars and active nuclei: 12th edition. *Astron. Astrophys.* **455**, 773–776 (2006).
33. George, M. et al. On active galactic nuclei as sources of ultra-high energy cosmic rays. *Mon. Not. R. Astron. Soc.* **388**, L59–L63 (2008).
34. Ghisellini, G. et al. Ultra-high energy cosmic rays, spiral galaxies and magnetars. Preprint at (<http://arxiv.org/abs/0806.2393>) (2008).
35. Abbasi, R. U. et al. Search for correlations between HiRes stereo events and active galactic nuclei. Preprint at (<http://arXiv.org/abs/0804.0382>) (2008).
36. LHC. 2008. (<http://lhc2008.web.cern.ch/lhc2008>) (2008).
37. Kawai, H. et al. Telescope array experiment. *Nucl. Phys. Proc.* **175–176** (Suppl.), 221–226 (2008).
38. Ebisuzaki, T. et al. The JEM-EUSO project: Observing extremely high energy cosmic rays and neutrinos from the International Space Station. *Nucl. Phys. Proc.* **175–176** (Suppl.), 237–240 (2008).
39. Orbiting wide-angle light-collectors (OWL). (<http://owl.gsfc.nasa.gov>) (2004).
40. Falcke, H. et al. Detection and imaging of atmospheric radio flashes from cosmic ray air showers. *Nature* **435**, 313–316 (2005).

Acknowledgements We express our gratitude to many people, too many to mention individually, in the AMANDA and Auger Collaborations, who guided us through the cosmic-ray world over the years. We especially thank our mentor A. Filevich, whom introduced us to cosmic-ray physics and the scientific world.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to P.M.B. (bauleo@amar.colostate.edu).

ARTICLES

Kinetochores geometry defined by cohesion within the centromere

Takeshi Sakuno^{1,2}, Kenji Tada^{1,3} & Yoshinori Watanabe^{1,3}

During cell division microtubules capture chromosomes by binding to the kinetochore assembled in the centromeric region of chromosomes. In mitosis sister chromatids are captured by microtubules emanating from both spindle poles, a process called bipolar attachment, whereas in meiosis I sisters are attached to microtubules originating from one spindle pole, called monopolar attachment. For determining chromosome orientation, kinetochore geometry or structure might be an important target of regulation. However, the molecular basis of this regulation has remained elusive. Here we show the link between kinetochore orientation and cohesion within the centromere in fission yeast *Schizosaccharomyces pombe* by strategies developed to visualize the concealed cohesion within the centromere, and to introduce artificial tethers that can influence kinetochore geometry. Our data imply that cohesion at the core centromere induces the mono-orientation of kinetochores whereas cohesion at the peri-centromeric region promotes bi-orientation. Our study may reveal a general mechanism for the geometric regulation of kinetochores, which collaborates with previously defined tension-dependent reorientation machinery.

A well-defined mechanism for the regulation for kinetochore orientation is the tension-dependent stabilization of kinetochore–microtubule attachment; unstable spindle microtubules repeatedly attach and release (reorient) kinetochores until they capture paired kinetochores from opposite sides, thus generating tension across cohered centromeres^{1,2}. During meiosis I, however, homologous chromosomes are connected by a reciprocal recombination (chiasmata) and, therefore, tension can be generated not only when sister chromatids are captured from opposite sides but also when homologous chromosomes are pulled from opposite sides^{1,3,4}. Indeed, meiotic cells always take the latter option at the first meiotic division. Importantly, even if recombination is genetically abolished during meiosis, sister kinetochores apparently refuse bipolar attachment, although tension-generating attachment can be established only in this way. Therefore, the structure or geometry of sister kinetochores may be modified to allow orientation towards the same surface at this stage of meiosis.

The geometric aspect of kinetochore orientation has long been recognized in vertebrates⁵. The staining of human interphase nuclei with anti-centromere antibodies revealed that the centromere is duplicated and resolved by the end of interphase⁶. This physical separation or resolution of sister centromeres would be important for back-to-back assembly of sister kinetochores, thus facilitating bipolar attachment to microtubules in mitosis. In contrast to mitosis, electron microscopic analyses of several animal germ cells have shown that sister kinetochores orient side-by-side and fuse in meiosis I^{3,7–9}. Despite its paramount importance for chromosome segregation, the molecular mechanism underlying the regulation of kinetochore geometry has remained largely elusive¹⁰.

Fission yeast centromeres, like metazoan centromeres, are composed of two domains, a kinetochore-assembling core centromere and heterochromatic peri-centromeric regions; a single kinetochore can attach several microtubules^{11,12}. One plausible model for the geometric regulation of kinetochores arose from the analysis of the sister chromatid cohesion molecule (cohesin) in this organism¹³. Mitotic Rad21-containing cohesin localizes preferentially to the

peri-centromeric regions^{14–16}, whereas meiotic Rec8-containing cohesin localizes additionally to the core centromere¹³. Crucially, the abolishment of Rec8 only at the core centromere results in equational rather than reductional division at meiosis I, advocating a model whereby the establishment of cohesion at the core centromere conjoins the two kinetochore domains at meiosis I, whereas the core regions may open to opposite sides when not establishing this cohesion¹⁷. Fission yeast Moa1, a meiosis-specific kinetochore protein, also has an essential role in establishing the mono-orientation of kinetochores, whereas centromeric Rec8 localization increases, rather than decreases, in *moa1Δ* cells¹⁷. Moreover, the contribution of cohesin to kinetochore orientation has been disputed in budding yeast^{18,19}. Thus, the lack of observation of actual cohesion at the core centromere in any organism leaves the cohesion-mediated mono-orientation model largely hypothetical.

Visualization of core centromere cohesion in meiosis

To visualize the cohesion of DNA sequences lying under kinetochores, we inserted a *lac* operator (*lacO*) array into the centre of the core centromere and expressed green fluorescent protein–lactose repressor (GFP–LacI) fusion proteins (*cnt1*–GFP) within the cell (Fig. 1a). The engineered centromere was shown to preserve intact function (Supplementary Fig. 1). However, this fluorescence marker might not reveal the cohesion state at the core centromere in the context of the chromosome because of the intact cohesion at neighbouring sites. To overcome this problem, we excised the DNA from the neighbouring chromosomal domains by a site-specific recombinase (R recombinase of *Zygosaccharomyces rouxii*) and circularized it^{20,21} (Fig. 1a). Given that cohesion is usually established during S phase, we excised the DNA region during post-DNA replication in meiosis by expressing R recombinase with a meiosis-specific *spo5*⁺ promoter¹³. A Southern blot assay confirmed the proper excision of the DNA duplex as a circle in zygotes (Fig. 1a). To visualize the uncoupling of the core centromere marked with GFP signals, we coloured the outer flanking *dh1L* site differently with the insertion

¹Laboratory of Chromosome Dynamics, Institute of Molecular and Cellular Biosciences, ²Promotion of Independence for Young Investigators, ³Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Yayoi, Tokyo 113-0032, Japan.

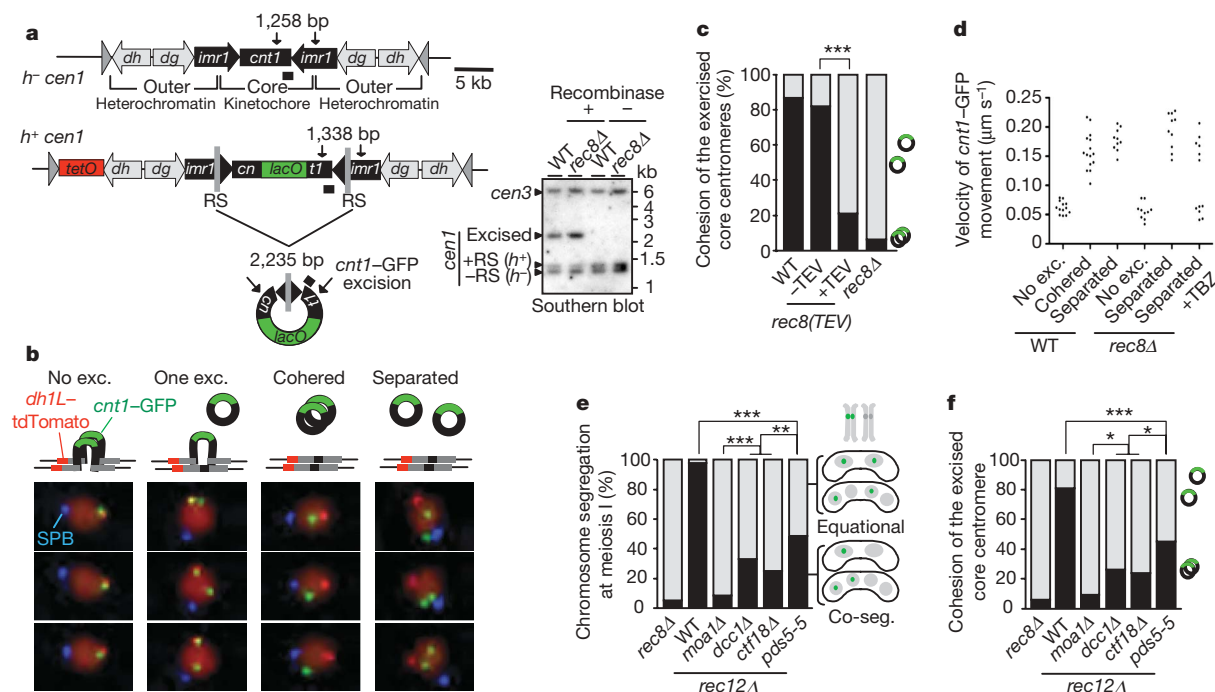


Figure 1 | Visualization of centromeric cohesion by excision from the chromosome. **a**, Schematic representation of centromere 1, in which a central core (*cnt1*) is surrounded by innermost repeats (*imr1*) and outer repeats (*dg* and *dh*), and its engineered version designed for excision. A split of *cnt1* by an insertion is designated *cn* and *t1*. RS, recombination sequence. Arrows indicate the positions of NdeI site and the length of fragments hybridized by the DNA probe (marked by the small filled box). DNA prepared from cells induced to meiotic prophase arrest was digested by NdeI and hybridized with the probe in the *cnt1* region (right bottom). Note that more than half of cells do not mate so that excision does not reach 50% at its maximum. Centromere 3 (*cen3*) was also cross-hybridized by the probe (77% identity). **b**, Centromere excision at prophase I was filmed and classified into four categories ('no exc.', not excised at all; 'one exc.', only one centromere is excised; 'cohered', both are excised with cohesion preserved; 'separated', both are excised without cohesion). The representative

of a *tet* operator (*tetO*) array and the expression of *tet* repressor–tdTomato fusion proteins (*dh1L*–tdTomato) (Fig. 1a). Filming wild-type prophase I cells revealed that *cnt1*–GFP signals were indeed separated from the *dh1L*–tdTomato dot, indicative of the excision of sister centromeres (Fig. 1b and Supplementary Videos 1–4), and most (>80%) of the excised centromeres preserved cohesion (Fig. 1c and Supplementary Fig. 2). We next applied this strategy to *rec8(TEV)* cells, in which endogenous Rec8 is mutated to become cleavable by tobacco etch virus (TEV) protease, and thus can be inactivated only around core centromeres by co-expressing Cen–TEV (TEV protease fused with a kinetochore-localizing peptide); otherwise, the Rec8 function remains intact¹⁷. When Cen–TEV was not expressed, the excised core centromeres largely preserved cohesion, as in wild-type *rec8*⁺ cells. In contrast, *rec8(TEV)* cells expressing Cen–TEV exhibited extensive separation (~80%) of the excised core centromeres (Fig. 1c). The separation reached >90% in *rec8Δ* cells, in which Rad21—including cohesin instead localizes to centromeres, thereby usually establishing bi-orientation²². These data imply that cohesion of the core centromere is established and preserved during prophase I, largely dependent on Rec8—including cohesin, a factor required for mono-orientation²³.

Filming the prophase I nucleus by time-lapse microscopy revealed that the excised kinetochore including *cnt1*–GFP dots moved actively depending on nuclear microtubules (Fig. 1b and Supplementary Fig. 3). Measurements of the movement indicated that both cohered centromeres in wild-type cells and separated centromeres in *rec8Δ* cells are similarly active (Fig. 1d). Notably, even when sister centromeres

were accidentally separated in wild-type cells, the movement of both centromeres was always active (Fig. 1d, WT separated). Thus, in prophase I, sister kinetochores destined for monopolar attachment are both active in association with microtubules, arguing against the hypothesis that the inactivation of one of the sister kinetochores leads to monopolar attachment at meiosis I. Instead, our results are consistent with the notion that both sister kinetochores are capable of attaching to microtubules, but the geometric restriction enforced by cohesion at the core centromeres promotes sister kinetochores to face the same side at meiosis I.

Core centromere cohesion links to mono-orientation

We assayed core centromere cohesion in several mutants defective in mono-orientation, including general cohesion factors, Ctf18, Dcc1 and Pds5, and the meiosis-I-specific kinetochore protein Moa1, which were all identified in a previous genetic screening¹⁷. The defect in mono-orientation is mildest in the *pds5-5* mutation (an insertion of the *ura4*⁺ cassette in the 5'-untranslated region), intermediate in *ctf18Δ* and *dcc1Δ*, and most extensive in *moa1Δ*, as in *rec8Δ*¹⁷ (Fig. 1e). Despite the fact that centromeric Rec8 localization in *moa1Δ* cells is even more abundant than in wild-type cells¹⁷, *moa1Δ* cells exhibit separation of the excised core centromeres as extensively as in *rec8Δ* cells (Fig. 1f). This result verifies the previous key prediction that Moa1 is required for the establishment and/or maintenance of cohesion at the core centromere¹⁷. In contrast, the cohesion defect was mildest in *pds5-5* cells and intermediate in *ctf18Δ* and *dcc1Δ* cells, revealing a remarkable quantitative correlation

between mono-orientation and core centromere cohesion (Fig. 1e, f). If cohesion at the core centromere leads to the mono-orientation of sister kinetochores, this cohesion should be disrupted before meiosis II, where bi-orientation is predominant. Indeed, when R recombinase was induced in zygotes arrested at prophase II, the excised core centromeres were largely separated, whereas overall centromeric cohesion was preserved (Supplementary Fig. 4). These results provide the strongest evidence that cohesion at the core centromere is closely linked to the mono-orientation of kinetochores at least during meiosis.

Lack of core centromere cohesion in mitosis

We next examined the mitotic centromeres. In the live imaging of proliferating cells, we did not detect any splitting of the *cnt1*-GFP signals derived from duplicated centromeres during G2 phase, although they frequently separated along the metaphase spindle (Supplementary Fig. 5a). Further measurement of the distance between sister centromeres at several sites, including the outer boundary of the core centromere (*imr1*-GFP) and at the outside of the centromere (~5 kb apart from the peri-centromeric repeats) on chromosome 2 (*cen2*-GFP), indicated that separation is most extensive at *cnt1*-GFP, and less at *imr1*-GFP, *dh1*-GFP and *cen2*-GFP in this order, consistent with the predicted geometry of centromeres under the tension of bipolar attachment (Fig. 2a, b and Supplementary Fig. 5b). The transient splitting of centromeres before anaphase may correspond to 'centromere breathing', which is well documented in budding yeast as well as in mammals^{6,24–27}. However, this breathing is generally thought to be the result of the overwhelming force of the spindle against intact cohesion, rather than being caused by the intrinsic loss of cohesion at the centromere. We challenged this view by applying the centromere excision at the core centromere (*cnt1*-GFP) and peri-centromeric region (*imr1*-GFP) (Figs 1b and 2c–e). The assays indicate that the core centromere largely separated after the excision induced during G2 phase, whereas peri-centromeric repeats retained cohesion in the same assay (Fig. 2f). The localization of the cohesin complex including Rad21 less to the core centromere than to the peri-centromeric repeats might account for the observed difference in cohesion¹³ (Fig. 2h). However, the ectopic expression of Rec8 in place of Rad21 establishes cohesion at the peri-centromeric repeats but still not at the core centromere (Fig. 2g), despite sufficient cohesin enriched at the core centromere (Fig. 2h). We therefore conclude that, regardless of the localization of cohesin, the establishment of cohesion is intrinsically avoided at the core centromere during mitosis.

An artificial tether restores meiotic mono-orientation

If conjoining at the core centromere is a mechanism to define the kinetochore geometry of mono-orientation, any kind of tether at this region should promote mono-orientation. To test this possibility, we made an artificial tether by using a tight endosomal adaptor/scaffold complex consisting of MP1 and p14 from mouse (refs 28 and 29). We fused these proteins with LacI, expecting that these proteins would hold two DNA duplexes together because the dimers potentially bind to two *lacO* sequences (Fig. 3a). To test this assumption, we expressed MP1-LacI and p14-LacI in *moa1Δ* cells, which carry a *lacO* array at *cnt1*, and examined the cohesion of the core centromere after excision (note that all MP1 and p14 peptides were fused by GFP and therefore act as fluorescence markers of centromeres). As a reference, we expressed MP1 instead of its LacI-fused version (Fig. 3a). As predicted, the expression of MP1-LacI/p14-LacI, but not MP1/p14-LacI, restored the cohesion of the excised core centromeres in *moa1Δ* cells (Fig. 3b).

We then applied this artificial tether to *moa1Δ rec12Δ* cells, in which both core centromere cohesion and mono-orientation at meiosis I are largely defective. When non-tethering MP1/p14-LacI was expressed, bi-orientation was predominant (>90% equational segregation; Fig. 3c, *moa1Δ rec12Δ*, *cnt1*, tether–). Similarly, MP1/p14-LacI did

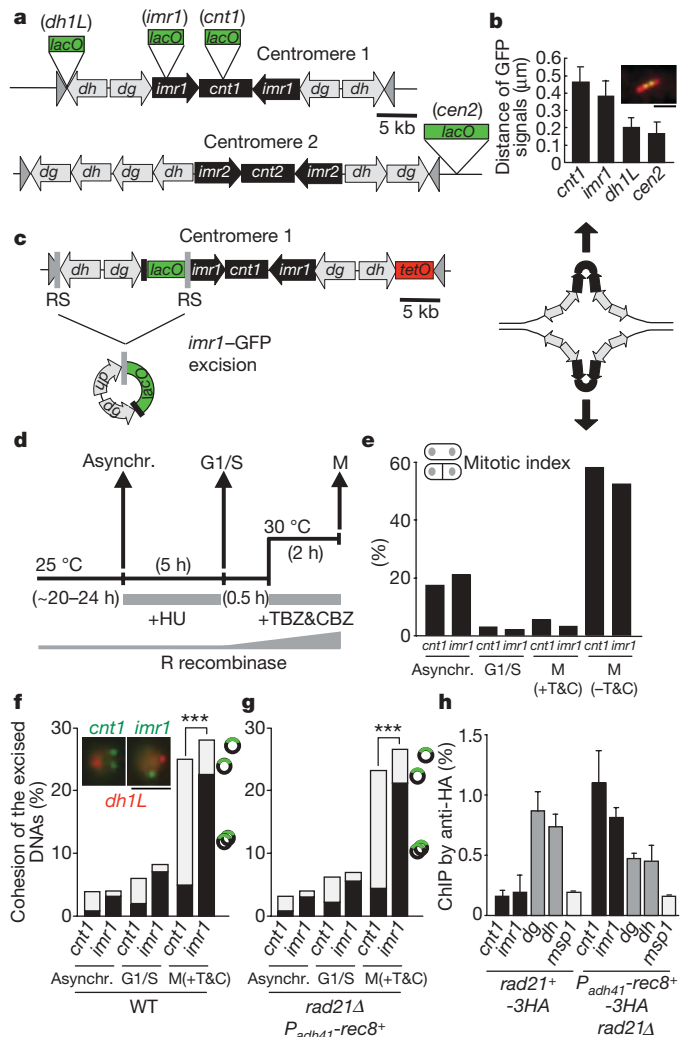


Figure 2 | Cohesion is avoided at the core centromere in mitosis.

a, Schematic illustration of centromeres of chromosomes 1 and 2, including positions used for fluorescent labelling. **b**, Average of the distance of split GFP signals at *cnt1*, *imr1L*, *dh1L* and *cen2* obtained by time-lapse imaging during metaphase (Supplementary Fig. 5b). Error bars, s.d. ($n = 10$). The predicted configuration of the metaphase centromere is depicted. **c**, Scheme for the excision of the peri-centromeric repeats. **d**, Schematic overview of the experiments. Asynchronous (asynch.) proliferating cells engineered at the centromere were synchronized once at G1/S phase by the addition of hydroxyurea (HU), and then released into medium containing spindle poisons (TBZ and carbendazim (CBZ)) so that they subsequently synchronized at prometaphase (M). As the expression of R recombinase is driven by the thiamine-repressible *nmt1*⁺ promoter, the duration of thiamine depletion was adjusted to excise the centromere mostly after the release from G1/S and during prometaphase arrest. **e**, Population of mitotic cells was quantified at each point in *cnt1*-GFP and *imr1*-GFP cells. As a reference for cell cycle progression, cells without TBZ and CBZ (T&C) were also prepared and analysed. **f**, **g**, GFP signals derived from excised *cnt1*-GFP and *imr1L*-GFP were examined in wild-type cells (**f**) and *rad21Δ* *P_{adh41}-rec8*⁺ cells (**g**). Only cells with both sister centromeres excised were counted. **h**, Proliferating cells were examined for the association of cohesin to the centromeric (*cnt1*, *imr1*, *dg* and *dh*) and arm (*msp1*) regions by chromatin immunoprecipitation (ChIP) assay using anti-hemagglutinin (HA) antibody. Error bars, s.d. ($n = 3$). Scale bars, 2 μ m.

not impair mono-orientation or co-segregation of sisters in *moa1*⁺ *rec12Δ* cells, indicating that the localization of the MP1/p14 complex itself does not perturb the intrinsic kinetochore function in general. In contrast, the expression of tethering MP1-LacI/p14-LacI restored mono-orientation in ~60% of *moa1Δ rec12Δ* cells (Fig. 3c, *moa1Δ rec12Δ*, *cnt1*, tether+). By simultaneously marking centromere 2, we confirmed that co-segregation occurred selectively in the

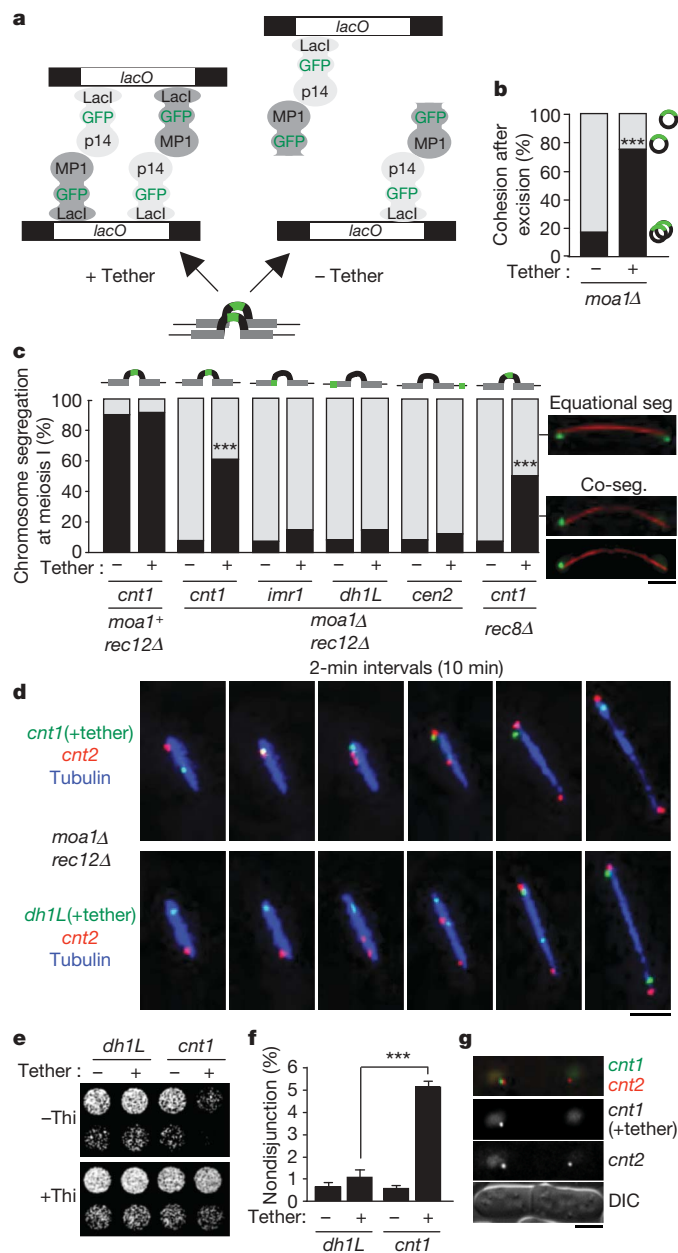


Figure 3 | Artificial tether in the core centromere restores mono-orientation. **a**, Scheme for tethering *lacO* array duplexes. **b**, Capacity of p14/MP1 tethering at *cnt1* was examined in *moa1Δ* cells as described in the legend to Fig. 1 ($n > 110$, $***P < 0.001$). **c**, Segregation of chromosomes tethered at various centromeric sites was examined at meiosis I in the indicated cells ($n > 140$, $***P < 0.001$). **d**, Time-lapse images of meiosis I were collected for 10 min at 2-min intervals in *moa1Δ rec12Δ* cells. Note that the upper cell is undergoing nondisjunction of tethered *cnt1*, whereas the lower cell is undergoing disjunction of tethered *dh1L*, with *cnt2*-tdTomato marked in different chromosomes is undergoing disjunction in either cells. **e–g**, Effect of tethering at centromeres in mitotic cells. **e**, Serial dilutions of the indicated cells were plated on MM medium with or without thiamine (Thi), which represses the expression of tethering proteins, and the cells were incubated at 28 °C for 3 days. **f**, Nondisjunction of tethered chromosomes was examined at 26.5 °C ($n > 400$). Error bars, s.d. ($n = 3$). $***P < 0.001$. **g**, Cells exhibiting nondisjunction of *cnt1*-GFP (tethered) showed disjunction of *cnt2*-tdTomato (non-tethered) ($n = 30$). DIC, differential interference contrast. Scale bars, 2 μ m.

tethered centromere 1, but never in the non-tethered centromere 2, even in the same cell ($n = 13$), indicating the specificity of the restored mono-orientation (Fig. 3d). For references, we applied an artificial tether at several sites within or in the vicinity of the centromere

(*imr1*, *dh1L* and *cen2*; see Fig. 2a and Supplementary Fig. 6). Consequently, tethering at a site other than *cnt1* hardly impaired the equational segregation of sisters (Fig. 3c, d). These results imply that the artificial tethers applied here do not provide enough cohesion to override the opposing force of spindle microtubules (also see Supplementary Fig. 7). Finally, the tether at *cnt1* also promoted the co-segregation of sisters in *rec8Δ* cells (Fig. 3c, *rec8Δ*). Collectively, we conclude that an artificial tether at the core centromere can bypass the requirement for Moa1 or Rec8 and formulate the kinetochore geometry that promotes mono-orientation.

Artificial mono-orientation in mitosis

To test whether kinetochore orientation is governed by the same principle in mitosis, we applied the artificial tether to proliferating cells. We confirmed that the artificial tether could act in mitotic cells by the excision assay, and kinetochore assembly itself does not change by the tethering at the core centromere (*cnt1*; Supplementary Figs 8 and 9). Remarkably, the tethering at *cnt1* caused growth retardation with increasing nondisjunction of chromosome 1 at anaphase (Fig. 3e–g). These defects were nearly undetectable in cells tethered at the peri-centromeric region (*dh1L*). It is notable that co-segregation of sisters imposed by the core centromere tether is only $< 10\%$ during mitosis. However, live image analysis revealed that more than 40% of cells persist with monopolar attachment throughout normal metaphase, although they are largely reoriented depending on the aurora B kinase complex with a delayed onset of anaphase (Supplementary Fig. 10). Collectively, these data indicate that, although the activity of reorientation tends to override the geometric restriction of kinetochores, the artificial tether at the core centromere indeed produces a bias to mono-orientation at metaphase even during mitosis.

Peri-centromeric cohesion promotes bi-orientation

In fission yeast, the heterochromatin protein Swi6 recruits cohesin to the peri-centromeric regions, and *swi6Δ* cells as well as cohesin mutant cells show impaired centromeric cohesion and merotelic attachment (one kinetochore attached to microtubules from both poles) at mitosis^{15,16}. Because *swi6Δ* cells still exhibit centromere breathing, an indicator of tension (Supplementary Fig. 11), we envisaged that loosening cohesion in the peri-centromeric regions might disrupt the kinetochore geometry of back-to-back rather than tension across centromeres. Accordingly, the tether in the peri-centromeric regions (*dh1L*), but not the tether at the core centromere (*cnt1*), suppressed mis-segregation of *swi6Δ* cells (Fig. 4a), albeit the applied tether at *dh1L* might not produce tension (Supplementary Fig. 7). Moreover, the presence of a lagging chromosome caused by merotelic attachment was also suppressed by the peri-centromeric tether (Fig. 4a).

To delineate the contribution of centromeric cohesion to kinetochore geometry, we sought to remove cohesion along the entire chromosome length except for the artificial tether in the vicinity of the centromeres. In meiosis, arm cohesion is removed during anaphase I, but only centromeric cohesion is retained until meiosis II through shugoshin (Sgo1)-dependent protection. Therefore, the deletion of *sgo1*⁺ enables the removal of all cohesion between sister chromatids during meiosis I, leading to complete random segregation in ensuing meiosis II^{30,31} (Fig. 4b). Moreover, the spindle checkpoint is less robust in meiosis II than in other phases³², and thus is advantageous for assaying the orientation of chromosomes that might not generate tension (Supplementary Fig. 12). We then applied the MP1/p14 tether to this *sgo1Δ* meiosis, assayed the chromosome segregation in meiosis II and calculated the orientation indices, defined as percentage difference between bi-orientation and mono-orientation (Fig. 4b, c). We confirmed that a tether at *cnt1*, *imr1*, *dh1L*, *cen2* or *lys1* (arm site ~ 10 kb distant from the boundary of centromere 1) does not perturb the intrinsic mono-orientation and reductional segregation at meiosis I (Supplementary Fig. 13). The assays in meiosis II indicated that tethering at *imr1*, *dh1L* or *cen2* (all

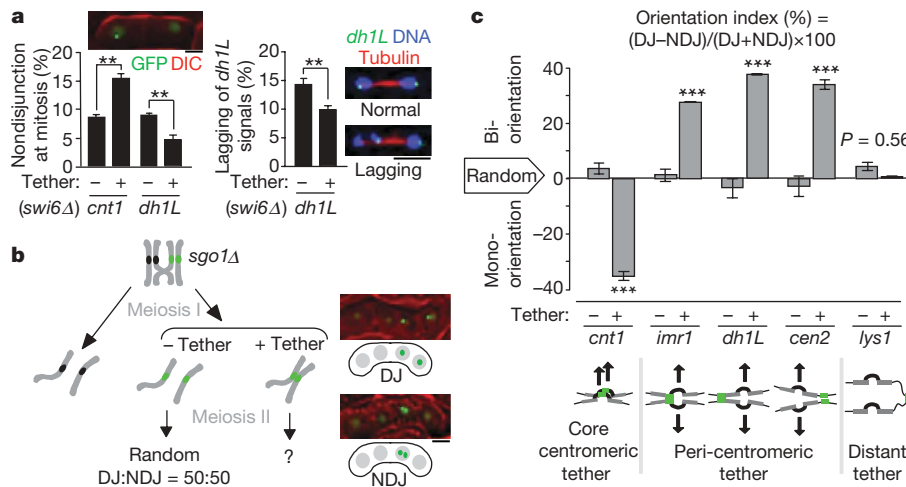


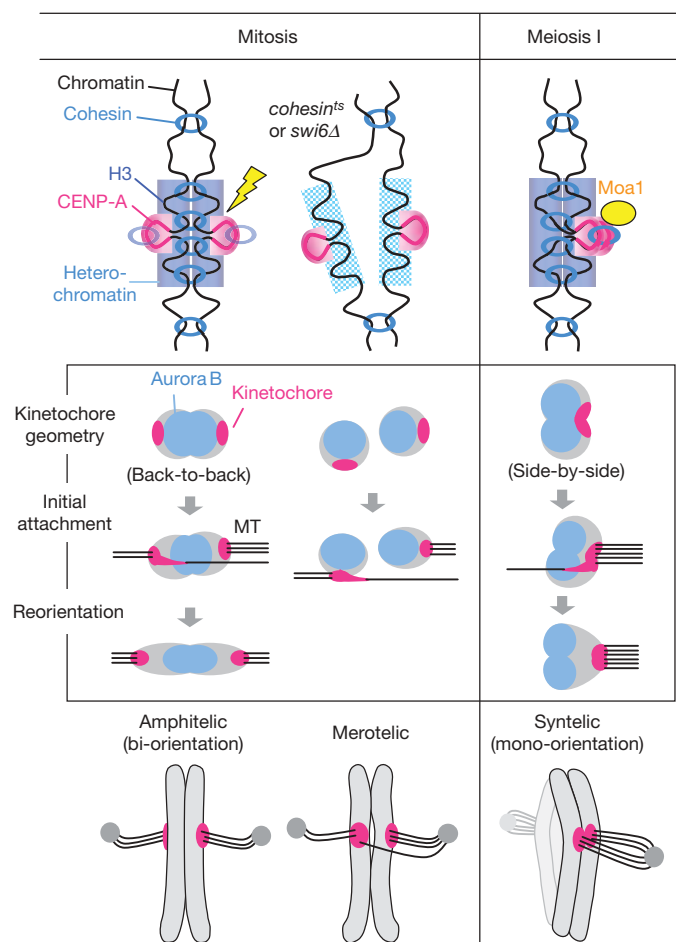
Figure 4 | Opposite effects of tethers at the core centromere and peri-centromeric regions. **a**, Frequency of nondisjunction ($n > 400$) and lagging chromosomes ($n > 150$) during mitosis were examined in the indicated *swi6Δ* cells at 18 °C. Error bars, s.d. ($n = 3$). Note that the tethering at *dh1L* suppresses both defects in *swi6Δ*. **b**, Schematic overview of the experiments

in the peri-centromeric regions) led to a bias towards bi-orientation (~ 30 – 40% preferential versus ~ 60 – 70% random segregation), whereas non-tethering MP1/p14 localization preserved randomness (Fig. 4c). In contrast, tethering at *cnt1* (in the core centromere) reversed the preference and caused a bias towards mono-orientation to a similar extent (Fig. 4c). This symmetric opposite bias of orientation between core and peri-centromeres argues that tethers applied in the vicinity of the centromere act through kinetochore geometry rather than tension that would uniformly bias towards bi-orientation. The fact that tethering at an arm site (*lys1*) near the centromere had little effect further supports this assumption (Fig. 4c). Collectively, these results indicate that the cohesion at the peri-centromeric regions facilitates a back-to-back configuration of kinetochores, whereas cohesion at the core centromere promotes a side-by-side configuration.

Discussion

Centromeric cohesion or the function of cohesin is well recognized as a mechanism to produce the tension across centromeres that plays a key part in stabilizing the connection of kinetochores to spindle microtubules and silencing the spindle checkpoint^{1,2}. However, the contribution of centromeric cohesion to kinetochore geometry has been largely unrecognized or indirectly speculated^{10,25,33–37}. Here we demonstrate that cohesion at the core centromere is established and maintained particularly during meiosis I (Fig. 1), but this cohesion is prevented during mitosis (Fig. 2). Ectopic localization of cohesin to the core centromere does not override this regulation in mitosis, which is reminiscent of *Moa1*-defective meiosis I, in which cohesin is nonfunctional at the core centromere despite its sufficient localization¹⁷. Thus, some mechanisms required for establishing cohesion might be intrinsically suppressed in the CENP-A (centromere-specific histone H3 variant) nucleosome region, and meiosis-I-specific kinetochore factors such as *Moa1* reverse this negative regulation on cohesion to promote mono-orientation (Fig. 5). An artificial tether at the core centromere efficiently restores mono-orientation to mutants in which cohesion at the core centromere is inherently deficient (Fig. 3). These results prove that mono-orientation of kinetochores is promoted ultimately by conjoining DNA duplexes underlying the kinetochores rather than by the action of a kinetochore protein itself. Moreover, although peri-centromeric cohesion might be important for generating tension across centromeres, our analyses reveal that this cohesion has an important role in promoting a back-to-back configuration of kinetochores (Fig. 4).

tethering centromeres in *sgo1Δ* meiosis. Examples of disjunction (DJ) and nondisjunction (NDJ) at meiosis II are shown at the right. **c**, Orientation indices for chromosomes tethered at various regions in the vicinity of centromeres (see Fig. 2a) were measured during meiosis II in *sgo1Δ* cells ($n > 240$). Error bars, s.d. ($n = 2$). ** $P < 0.01$; *** $P < 0.001$. Scale bars, 2 μ m.



We propose a model in which cohesion within the centromere regulates kinetochore geometry (Fig. 5). In fission yeast, the central core region of the centromere is likely to form inverted configurations that protrude from the peri-centromeric heterochromatic regions¹¹. Thus, the kinetochore assembled on chromatin is asymmetric along the chromosome axis and locates on the opposite side of heterochromatin, as observed in animal kinetochores^{10,38,39} (Fig. 5). Peri-centromeric heterochromatin would be a semi-compressible entity occupying three-dimensional space. Cohesion within this region tends to cause folding into a bulky configuration that would prevent the co-orientation of the kinetochores unless DNA duplexes under the kinetochores are specifically tethered by cohesin (or an artificial tether) as in meiosis I (Fig. 5). Moreover, kinetochore geometry would define the relative location between microtubule-attachment sites and 'reorientation factors' such as the aurora B kinase complex, which largely localizes at the peri-centromeric region (inner centromere in metazoa) and has a role in destabilizing the attachment^{40–42}. The cohesion at the peri-centromeric region involuntarily positions the reorientation factors between sister kinetochores, whereas cohesion at the core centromere would exclude the factors from locating between kinetochores, pushing them instead towards the opposite side (Fig. 5, kinetochore geometry). In either case, the attachment sites connected to the incorrect pole would be positioned close to the reorientation factors and, thus, be efficiently corrected (Fig. 5, reorientation). However, dissociation of sister chromatids around the centromere including the peri-centromeric regions, as observed in heterochromatin- or cohesin-defective mitotic cells, would result in a loss of the rotational restriction of the centromeric chromatin that otherwise brings the incorrect attachment sites close to the reorientation factors, thus rendering merotelic attachment persistent (Fig. 5). We therefore suggest that the geometry of kinetochores influences both initial attachment and reorientation.

Linear and three-dimensional observations of the centromeric structure of animal mitotic chromosomes indicate that CENP-A-containing nucleosomes line up on the poleward face, whereas the heterochromatic H3 nucleosome region may establish cohesion and form an internal surface^{34,35,38}. Moreover, sister kinetochores or CENP-A regions closely associate during meiosis I in germ cells⁸. These observations fit our results in fission yeast. The budding yeast 'point' centromere, which is composed of a single CENP-A nucleosome without surrounding heterochromatin, may also not be an exception, because cohesion at the core centromere is probably established only at meiosis I depending on the specific protein complex, monopolin^{18,19}. However, it is proposed that this cohesion does not require Rec8 or even cohesin, and that the jointed centromeres assemble a single kinetochore rather than side-by-side sister kinetochores at meiosis I^{19,43}. In contrast, maize and *Arabidopsis* Rec8 has an essential role in mono-orientation, as in fission yeast^{44,45}. Thus, the cohesion-mediated kinetochore geometry model revealed in fission yeast would be well conserved in metazoa and plants that carry the 'regional' centromere. Although further studies are required to formulate the conservation, our work in fission yeast provides a fundamental model for considering the regulation of kinetochore geometry in eukaryotic chromosomes.

METHODS SUMMARY

All *S. pombe* strains used in this study are listed in Supplementary Table 1. For the introduction of a *lacO* or *tetO* array into various centromeric regions, approximately 5.5 kb of a *lacO* array or 10 kb of a *tetO* array were used in this study, except for the previously constructed *cen2-lacO* and *lys1-lacO*^{6,47}. To prepare meiotic cells arrested at prophase I, logarithmic growing *h⁺* and *h⁻* *mei4Δ* cells⁴⁸ were collected, mixed, spotted on a SPA plate and incubated at 30 °C for 12 h. Similarly, we used *mes1Δ* cells⁴⁹ to arrest meiosis at prophase II. For time-lapse imaging of mitotic cells, logarithmically growing cells cultured in MM-Thi medium at 30 °C for at least 15 h were used. For time-lapse imaging of meiosis I progression, zygotes spotted on SPA plates at 26.5 °C for 7 h were used. Live-cell recordings were performed on a DeltaVision RT system (Applied Precision) in an air-conditioned room maintained at 26.5 °C. To measure the velocity of *cnt1*-GFP

signal movement, time-lapse observation at 3-s intervals was performed with *dh1L*-tdTomato. After a consecutive 90 s (30 time points) measurement, the total distance moved was divided by 45 s to obtain the velocity. The chromatin immunoprecipitation assay was carried out essentially as described previously¹⁷.

Statistics. Unpaired two-tailed *t*-tests were used for Supplementary Fig. 10c and Figs 3f and 4a; otherwise, Chi-squared tests were used.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 September 2008; accepted 12 February 2009.

- Nicklas, R. B. How cells get the right chromosomes. *Science* **275**, 632–637 (1997).
- Tanaka, T. U. Bi-orienting chromosomes on the mitotic spindle. *Curr. Opin. Cell Biol.* **14**, 365–371 (2002).
- Moore, D. P. & Orr-Weaver, T. L. Chromosome segregation during meiosis: building an unambivalent bivalent. *Curr. Top. Dev. Biol.* **37**, 263–299 (1998).
- Petronczki, M., Siomos, M. F. & Nasmyth, K. Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. *Cell* **112**, 423–440 (2003).
- Östergren, G. The mechanism of co-orientation in bivalents and multivalents. *Hereditas* **37**, 85–156 (1951).
- Brenner, S., Pepper, D., Berns, M. W., Tan, E. & Brinkley, B. R. Kinetochore structure, duplication, and distribution in mammalian cells: analysis by human autoantibodies from scleroderma patients. *J. Cell Biol.* **91**, 95–102 (1981).
- Goldstein, L. S. Kinetochore structure and its role in chromosome orientation during the first meiotic division in male *D. melanogaster*. *Cell* **25**, 591–602 (1981).
- Parra, M. T. et al. Involvement of the cohesin Rad21 and SCP3 in monopolar attachment of sister kinetochores during mouse meiosis I. *J. Cell Sci.* **117**, 1221–1234 (2004).
- Lee, J. et al. Specific regulation of CENP-E and kinetochores during meiosis I/meiosis II transition in pig oocytes. *Mol. Reprod. Dev.* **56**, 51–62 (2000).
- Hauf, S. & Watanabe, Y. Kinetochore orientation in mitosis and meiosis. *Cell* **119**, 317–327 (2004).
- Takahashi, K. et al. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol. Biol. Cell* **3**, 819–835 (1992).
- Pidoux, A. & Allshire, R. Kinetochore and heterochromatin domains of the fission yeast centromere. *Chromosome Res.* **12**, 521–534 (2004).
- Watanabe, Y., Yokobayashi, S., Yamamoto, M. & Nurse, P. Pre-meiotic S phase is linked to reductional chromosome segregation and recombination. *Nature* **409**, 359–363 (2001).
- Tomonaga, T. et al. Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14**, 2757–2770 (2000).
- Bernard, P. et al. Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542 (2001).
- Nonaka, N. et al. Recruitment of cohesin to heterochromatic regions by Swi6/HPI in fission yeast. *Nature Cell Biol.* **4**, 89–93 (2002).
- Yokobayashi, S. & Watanabe, Y. The kinetochore protein Moa1 enables cohesion-mediated monopolar attachment at meiosis I. *Cell* **123**, 803–817 (2005).
- Toth, A. et al. Functional genomics identifies monopolin: a kinetochore protein required for segregation of homologs during meiosis I. *Cell* **103**, 1155–1168 (2000).
- Monje-Casas, F., Prabhu, V. R., Lee, B. H., Boselli, M. & Amon, A. Kinetochore orientation during meiosis is controlled by Aurora B and the monopolin complex. *Cell* **128**, 477–490 (2007).
- Araki, H. et al. Site-specific recombinase, R, encoded by yeast plasmid pSR1. *J. Mol. Biol.* **225**, 25–37 (1992).
- Chang, C. R., Wu, C. S., Hom, Y. & Gartenberg, M. R. Targeting of cohesin by transcriptionally silent chromatin. *Genes Dev.* **19**, 3031–3042 (2005).
- Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell Biol.* **23**, 3965–3973 (2003).
- Watanabe, Y. & Nurse, P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature* **400**, 461–464 (1999).
- Goshima, G. & Yanagida, M. Establishing biorientation occurs with precocious separation of sister kinetochores, but not the arms, in the early spindle of budding yeast. *Cell* **100**, 619–633 (2000).
- He, X., Asthana, S. & Sorger, P. K. Transient sister chromatid separation and elastic deformation of chromosomes during mitosis in budding yeast. *Cell* **101**, 763–775 (2000).
- Ocampo-Hafalla, M. T., Katou, Y., Shirahige, K. & Uhlmann, F. Displacement and re-accumulation of centromeric cohesin during transient pre-anaphase centromere splitting. *Chromosoma* **116**, 531–544 (2007).
- Rieder, C. L. & Salmon, E. D. Motile kinetochores and polar ejection forces dictate chromosome position on the vertebrate mitotic spindle. *J. Cell Biol.* **124**, 223–233 (1994).
- Kurzbaue, R. et al. Crystal structure of the p14/MP1 scaffolding complex: how a twin couple attaches mitogen-activated protein kinase signaling to late endosomes. *Proc. Natl Acad. Sci. USA* **101**, 10984–10989 (2004).
- Gruber, S. et al. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* **127**, 523–537 (2006).

30. Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* **427**, 510–517 (2004).
31. Rabitsch, K. P. *et al.* Two fission yeast homologs of *Drosophila* Mei-5332 are required for chromosome segregation during meiosis I and II. *Curr. Biol.* **14**, 287–301 (2004).
32. Yamamoto, A. *et al.* Spindle checkpoint activation at meiosis I advances anaphase II onset via meiosis-specific APC/C regulation. *J. Cell Biol.* **182**, 277–288 (2008).
33. Tanaka, T., Fuchs, J., Loidl, J. & Nasmyth, K. Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. *Nature Cell Biol.* **2**, 492–499 (2000).
34. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Struct. Mol. Biol.* **11**, 1076–1083 (2004).
35. Amor, D. J., Kalitsis, P., Sumer, H. & Choo, K. H. A. Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol.* **14**, 359–368 (2004).
36. Dewar, H., Tanaka, K., Nasmyth, K. & Tanaka, T. U. Tension between two kinetochores suffices for their bi-orientation on the mitotic spindle. *Nature* **428**, 93–97 (2004).
37. Eckert, C. A., Gravidahl, D. J. & Megee, P. C. The enhancement of pericentromeric cohesin association by conserved kinetochore components promotes high-fidelity chromosome segregation and is sensitive to microtubule-based tension. *Genes Dev.* **21**, 278–291 (2007).
38. Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002).
39. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores. From epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421 (2003).
40. Tanaka, T. U. *et al.* Evidence that the Ipl1–Sli15 (Aurora kinase–INCENP) complex promotes chromosome bi-orientation by altering kinetochore–spindle pole connections. *Cell* **108**, 317–329 (2002).
41. Cimini, D., Wan, X., Hirel, C. B. & Salmon, E. D. Aurora kinase promotes turnover of kinetochore microtubules to reduce chromosome segregation errors. *Curr. Biol.* **16**, 1711–1718 (2006).
42. Kawashima, S. A. *et al.* Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21**, 420–435 (2007).
43. Winey, M., Morgan, G. P., Straight, P. D., Giddings, T. H. Jr & Mastronarde, D. N. Three-dimensional ultrastructure of *Saccharomyces cerevisiae* meiotic spindles. *Mol. Biol. Cell* **16**, 1178–1188 (2005).
44. Yu, H.-G. & Dawe, R. K. Functional redundancy in the maize meiotic kinetochore. *J. Cell Biol.* **151**, 131–141 (2000).
45. Chelysheva, L. *et al.* AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. *J. Cell Sci.* **118**, 4621–4632 (2005).
46. Yamamoto, A. & Hiraoka, Y. Monopolar spindle attachment of sister chromatids is ensured by two distinct mechanisms at the first meiotic division in fission yeast. *EMBO J.* **22**, 2284–2296 (2003).
47. Nabeshima, K. *et al.* Dynamics of centromeres during metaphase–anaphase transition in fission yeast: *dis1* is implicated in force balance in metaphase bipolar spindle. *Mol. Biol. Cell* **9**, 3211–3225 (1998).
48. Horie, S. *et al.* The *Schizosaccharomyces pombe* *mei4⁺* gene encodes a meiosis-specific transcription factor containing a forkhead DNA-binding domain. *Mol. Cell Biol.* **18**, 2118–2129 (1998).
49. Izawa, D., Goto, M., Yamashita, A., Yamano, H. & Yamamoto, M. Fission yeast *Mes1p* ensures the onset of meiosis II by blocking degradation of cyclin Cdc13p. *Nature* **434**, 529–533 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hauf for critically reading the manuscript. We thank H. Matsuzaki, A. Yamamoto, R. Allshire, K. Nasmyth and the Yeast Genetic Resource Center (YGRC) for yeast strains. We also thank all the members of our laboratory for their support and discussion, especially S. Yokobayashi for materials and assistance in the initial stage of this project. This work was supported in part by Special Coordination Funds for Promoting Science and Technology (to T.S.), the Global COE Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms), MEXT, Japan, and a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to Y.W.).

Author Contributions T.S. and Y.W. conceived and designed the experiments. T.S. performed all experiments. K.T. set up the TetR–tdTomato system in fission yeast. Y.W. planned research and wrote the manuscript with input from T.S.

Author information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.W. ([ywatana@iam.u-tokyo.ac.jp](mailto:ywatanab@iam.u-tokyo.ac.jp)).

METHODS

S. pombe strain. All strains used in this study are described in Supplementary Table 1. Deletions of endogenous *rec12*⁺ by the hygromycin-resistant (*hyg*^r) or clonNAT-resistant (*nat*^r) marker, and *rad21*⁺ by the kanamycin-resistant (*kan*^r) marker, were performed using the PCR-based gene-targeting method for *S. pombe*. *P_{adh41}* (a weak version of the *adh1*⁺ promoter)-derived *rec8*⁺-3HA-*ura4*⁺ strain was created as described previously²². The construction of *rec8*(TEV) and mutant allele of *dcc1Δ*, *ctf18Δ* and *pds5-5* were described previously¹⁷. p14 and MP1 cDNAs were amplified from a mouse testis cDNA library (a gift from Y. Imai) by PCR. To express p14- or MP1-GFP-LacI in meiosis, a sequence encoding GFP and LacI containing nuclear localization signal (NLS) at its C terminus (in the case of MP1, a plasmid without LacI was also made) were fused to the C terminus of p14 or MP1 and cloned under the promoter *P_{nmt41}* or *P_{nmt1}*, respectively. For efficient termination of the transcripts, *T_{adh1}* was used. The resulting plasmids (pHBCN41-p14-GFP-LacI, pNATZN1-MP1-GFP+/−LacI) were linearized and integrated at the locus adjacent to the SPAC26F1.12c gene of chromosome 1 (we refer to this as the C locus) using the *hyg*^r marker, or the *zfs1*⁺ gene of chromosome 2 (we refer to this as the Z locus) using the *nat*^r marker. To express p14- and MP1-LacI simultaneously in proliferating cells, a *P_{nmt1}*-MP1-GFP-LacI-*T_{adh1}* or *P_{nmt1}*-MP1-GFP-*T_{adh1}* fragment was subcloned into pHBCN41-p14-GFP-LacI plasmid at the 3' end of its *T_{adh1}*. The resulting plasmid was linearized and integrated at the C locus using the *hyg*^r marker. The R recombinase coding region (R gene) was amplified from pNN115 (a gift from H. Matsuzaki, referred to in ref. 20) by PCR. To express R-CFP in mitotic phase, a sequence encoding CFP and the *T_{nmt1}* sequence from pREP1 were fused to the C terminus of the R gene cloned under the promoter *P_{nmt1}*. The resulting plasmid was linearized and integrated at the C locus using the *hyg*^r marker. To express R protein in meiotic prophase I (or II), the R gene was cloned between the meiosis-specific *spo5*⁺ (or *spo6*⁺) promoter (~500 bp) and the 3' untranslated region of *spo5*⁺ (~600 bp) that suppresses the basal mitotic expression. The resulting plasmid was linearized and integrated at the *lys1*⁺ locus of chromosome 1 using the *hyg*^r marker or at the Z locus using the *nat*^r marker. To visualize tubulin, besides the pREP81-CFP-*atb2*⁺ or pREP81-mCherry-*atb2*⁺ plasmid, *P_{adh15}* (a weak version of the *adh1*⁺ promoter)-CFP-*atb2*⁺ integrated at the C locus using the *hyg*^r marker or *P_{adh13}* (a bit stronger version of the *P_{adh15}*)-mCherry-*atb2*⁺ integrated at the Z locus using the *nat*^r marker was also used. To visualize the SPB, the C terminus of the endogenous *sad1*⁺ gene was tagged with CFP using the pUC119 plasmid containing *LEU2* cassette and a sequence encoding CFP fused with the C-terminal portion of the *sad1*⁺ gene. The *tetR*⁺ ORF was amplified from p128tetR-GFP (a gift from K. Nasmyth) by PCR. To visualize a *tetO* array integrated at the chromosome, a sequence encoding tdTomato containing NLS at its C terminus was fused with the C terminus of *tetR*⁺ ORF under the promoter *P_{adh31}* (a somewhat stronger version of *P_{adh41}*). The resulting plasmid was linearized and integrated at the Z locus using the *nat*^r marker.

Introduction of a *lacO* or *tetO* array into various centromeric regions. A *lacO* array engineered from pCT31 (a gift from A. Yamamoto) or a *tetO* array engineered from p306tetO₂x112 (a gift from K. Nasmyth) were used in this study. A fragment containing a *lacO* array with or without the *kan*^r cassette, flanked by the *cnt1* region (~3,769,353–3,770,866 bp and ~3,770,867–3,771,989 bp of chromosome 1), was transformed into *TM1* << *ura4*⁺ strain (derived from FY336, in which a *ura4*⁺ cassette was introduced at the NcoI site corresponding to ~3,770,864–3,770,869 bp of chromosome 1)³⁰. Uracil auxotroph colonies were selected on a FOA (5-fluoroorotic acid hydrate, Sigma) plate and correct integration was confirmed by PCR. Fragments containing *lacO* or *tetO* with the *ura4*⁺ cassette, flanked by the *imr1* regions (~3,764,007–3,764,631 bp and ~3,764,632–3,765,171 bp of chromosome 1), the *dh1L* regions (~3,751,647–3,753,648 bp and ~3,753,649–3,755,637 bp of chromosome 1) or the *cnt2* regions (~1,623,029–1,623,738 bp and ~1,623,739–1,624,423 bp of chromosome 2), were used for transformation to obtain *imr1-lacO*, *dh1L-lacO*/*tetO* or *cnt2-tetO*, respectively. The *dh1R-tetO* was obtained during the construct of *dh1L-tetO* because of the similarity in the surrounding sequences. Among *Ura*⁺ transformants, correct integration was selected and confirmed

by PCR. Among symmetric sequences of *imr1* or *dh1*, the integrated locus (left or right) was also determined by PCR.

Introduction of recombination sequences. Fifty-eight base pairs of recombination sequences (RSs)²⁰ was introduced into both the BamHI and SalI sites of pBluescript-SK⁺ vector containing a *ura4*⁺ cassette in a head-to-tail-orientation. The resultant plasmid pBS-RS-*ura4*⁺-RS was used as a template for the PCR-based gene targeting method. For *cnt1*-GFP excision, a RS-*ura4*⁺-RS-containing fragment was integrated on the left side of *cnt1* and subsequently excised by R recombinase, leaving single copy of RS at this locus (3,768,723 bp of chromosome 1, named FY534). Using this strain, a single copy of RS was further generated on the right side of *cnt1* (3,773,551 bp of chromosome I, FY527). For *imr1*-GFP excision, RSs were similarly introduced at the outer boundary of *dh1L* (3,752,623 bp of chromosome 1, FY937) and *imr1* (3,768,723 bp of chromosome 1, named FY501). Correct integration of RS was confirmed by PCR.

Centromere excision. For the centromeric excision assay in meiosis, cells of *h*⁺ *mei4Δ* carrying engineered chromosome 1 with RS-*cnt1*-GFP-RS and *dh1L*-tdTomato were crossed with *h*[−] *mei4Δ* with intact chromosome 1, and signals of *cnt1*-GFP and *dh1L*-tdTomato were observed in the prophase-I-arrested zygotes by time-lapse imaging at 17-s intervals for at least 6 min. For the centromeric excision assay in mitosis, cells were cultured in MM medium without thiamine (MM-Thi) at 25 °C for 20 h or 24 h, and then arrested at G1/S by the addition of hydroxyurea (Wako) at a final concentration of 12 mM; the cells were then cultured at 25 °C for a further 5 h. Thirty minutes after hydroxyurea removal, cells were suspended in MM-Thi medium containing DMSO (final 1%) or 50 μg ml^{−1} of TBZ (Sigma) and CBZ (Wako), and cultured at 30 °C for 2 h (to arrest cells at prometaphase).

Centromere tethering. For the analyses of centromere tethering in meiosis, cells were cultured in MM-Thi at 30 °C for 15 h to express MP1 or p14 before being spotted on an SPA plate. After ~9–12 h or ~12–15 h of spotting at 26.5 °C, meiosis I or meiosis II was observed, respectively. The general methods for the induction of meiosis and the monitoring of chromosome segregation were as described previously¹⁷.

Time-lapse imaging. For the measurement of anaphase onset in wild type or *bir1-T1*, cells were cultured in MM-Thi medium at 26.5 °C for 18 h, and then live-cell recordings were performed on a DeltaVision P system (Applied Precision) in an air-conditioned chamber maintained at 29 °C. A glass-bottom dish (Matsumi) coated with 0.2% Concavalin A (Sigma) was used to mount cells. During observation, the cells transferred to the glass-bottom dish were supplied with MM-Thi medium for mitosis, or MM medium without nitrogen (MM-N) for meiosis. For the observation of meiosis, zygotes suspended in MM-N were sonicated briefly before mounting onto the glass-bottom dish to avoid the aggregation. For the measurement of the velocity of *cnt1*-GFP signals, zygotes mounted onto the glass-bottom dish were supplied with MM-N containing 0.1% of DMSO or TBZ (50 μg ml^{−1}). Images were acquired using the Z-sweep acquisition (OAI) feature. In other cases, images were acquired by the Z-sectioning and stacked using 'quick projection' in the softWoRx software.

Southern hybridization. Genomic DNAs were prepared from cell cultures spotted on SPA plates and incubated at 30 °C for 14 h (mating efficiency was ~40% in all cultures). In total, 40 μg of DNA subjected to NdeI digestion was used for gel electrophoresis and blotting. A region from 3,772,687 bp to 3,773,186 bp of chromosome 1 was used as a probe DNA.

Chromatin immunoprecipitation assay. An anti-haemagglutinin monoclonal (12CA5, Sigma) antibody was used for immunoprecipitation. DNA prepared from whole-cell extracts or immunoprecipitated fractions was analysed by quantitative PCR with the ABI PRISM7000 system (Applied Biosystems) using SYBR Premix Ex Taq (Perfect Real Time) (Takara). The primers used for PCR were all described previously¹⁷. Immunoprecipitation by an untagged strain was accounted for as nonspecific binding in ChIP fractions.

50. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–233 (1995).

Design of protein-interaction specificity gives selective bZIP-binding peptides

Gevorg Grigoryan^{1†}, Aaron W. Reinke¹ & Amy E. Keating¹

Interaction specificity is a required feature of biological networks and a necessary characteristic of protein or small-molecule reagents and therapeutics. The ability to alter or inhibit protein interactions selectively would advance basic and applied molecular science. Assessing or modelling interaction specificity requires treating multiple competing complexes, which presents computational and experimental challenges. Here we present a computational framework for designing protein-interaction specificity and use it to identify specific peptide partners for human basic-region leucine zipper (bZIP) transcription factors. Protein microarrays were used to characterize designed, synthetic ligands for all but one of 20 bZIP families. The bZIP proteins share strong sequence and structural similarities and thus are challenging targets to bind specifically. Nevertheless, many of the designs, including examples that bind the oncoproteins c-Jun, c-Fos and c-Maf (also called JUN, FOS and MAF, respectively), were selective for their targets over all 19 other families. Collectively, the designs exhibit a wide range of interaction profiles and demonstrate that human bZIPs have only sparsely sampled the possible interaction space accessible to them. Our computational method provides a way to systematically analyse trade-offs between stability and specificity and is suitable for use with many types of structure-scoring functions; thus, it may prove broadly useful as a tool for protein design.

Designing peptides, proteins or small molecules that bind to native protein targets is a promising route to new reagents and therapies. However, dealing with the interaction specificity problem—that is, achieving designs that are selective for their intended targets in preference to related alternatives—is difficult. Designing or assessing protein-interaction specificity in a comprehensive manner is impeded by the challenges and costs inherent in modelling or measuring many competing complexes. Recent large-scale experiments that have characterized interaction specificity for a handful of protein families and/or domains represent significant progress in this area^{1–6}. In particular, assays that provide a way to profile the interactions of a protein with many candidate partners offer an opportunity to explore how specificity can be introduced into proteins rationally, by design.

Computational design has led to remarkable advances in protein engineering over the past decade, including the design of protein–protein interactions^{7–15}. Introducing considerations of specificity into protein-design calculations raises interesting theoretical challenges that have been addressed in a few previous studies^{7,16,17} and/or treated on a case-by-case basis in several applications^{7–10,15}. Most often, however, specificity is simply ignored in computational protein design. Several proteins or peptides that were optimized solely for binding to a native target were shown *a posteriori* to be specific for their intended interaction partner over a few related alternatives^{11–14}. However, focusing only on the stability of the desired complex led to a lack of specificity, both in computational design and experimental selections; in other examples^{15,16,18}. Strategies that can simultaneously consider affinity and multi-state specificity in the design process are therefore highly desirable⁷.

bZIP transcription factor coiled coils

The bZIP transcription factors provide an exciting but highly challenging opportunity to test strategies for interaction-specificity design. The bZIPs homo- and/or hetero-dimerize by forming a parallel coiled

coil (a ‘leucine zipper’) and bind DNA using a region rich in basic amino acids¹⁹. Approximately 53 human bZIP proteins that make up 20 families participate in a wide range of important biological processes and pose attractive targets for selective inhibition. Interest in inhibiting bZIPs dates to 1995, when it was shown that heterodimers containing one bZIP subunit and one subunit with an acidic region replacing the basic region (A-ZIPs) are inactive. A-ZIPs have proven very useful for applications both *in vitro* and *in vivo*^{20,21}. However, these inhibitors mimic the interaction preferences of the proteins from which they are derived and typically associate with multiple bZIP families. Extensive sequence similarity among the leucine-zipper domains hampers efforts to make specific peptides that could provide more selective A-ZIPs or other inhibitors. For example, strong undesirable off-target interactions were observed when experimentally selecting synthetic partners for the c-Fos and c-Jun bZIP coiled coils out of peptide libraries¹⁸.

The bZIPs are also attractive design targets because experiments have probed sequence features that influence both structural and interaction specificity^{19,22–24}. Building on these insights and taking advantage of large experimental data sets, computational models that provide useful predictions of bZIP interaction preferences have been developed^{4,18,25,26}. These previous studies afford a relatively mature understanding of bZIP partnering and provide the potential for specificity design.

Computational design of specificity

We have developed a strategy for addressing specificity in protein-design calculations that rests on the trade-off between maximizing affinity and introducing specificity. The stability/specificity trade-off has been discussed previously^{7,15–17}, and has motivated the successful design of heterospecific coiled-coil pairs⁷. For our work, we note that a protein designed to bind optimally to a native target may also bind strongly to one or more undesired competitor, indicating that the difference in energy between forming undesired complexes and the

¹MIT Department of Biology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. †Current address: University of Pennsylvania Department of Biochemistry and Biophysics, 422 Curie Blvd, Philadelphia, Pennsylvania 19107, USA.

design–target complex is not sufficiently large. New designs can be sought that increase this gap and are thus more selective for the target, but these will necessarily have reduced target affinities relative to the design that is optimal for target binding. The computational method presented here formalizes this trade-off by identifying sequences that minimize the stability sacrifice required to achieve increasing energy gaps from competing complexes. Sequences designed in this way cannot simultaneously be improved both in predicted affinity and in specificity.

Our framework, CLASSY (cluster expansion and linear programming-based analysis of specificity and stability), makes use of two computational techniques to implement the above-mentioned idea. The first is integer linear programming (ILP), an optimization method that has been applied to the energy-minimization problem in protein design²⁷. The second is cluster expansion, which we use to convert a structure-based interaction model into a sequence-based scoring function that is very fast to evaluate^{28,29}. Importantly, cluster expansion allows us to apply ILP at the sequence level, rather than at the structure level. This makes it possible to impose constraints on the energies of design–undesired partner interactions during optimization of the design–target energy, which is the keystone of the CLASSY approach. The power of cluster expansion and ILP means that arbitrary numbers of desired and undesired states and relationships between them can be included in CLASSY designs. Thus, CLASSY can deal with

problems beyond the scope of traditional design methods, making it an appropriate approach for designing specific anti-bZIP peptides.

As one example of how CLASSY can be used, we implemented a procedure called a ‘specificity sweep’ to identify sequences of optimal stability that satisfy increasing requirements on specificity. For this purpose, the quantity Δ was defined as the energy gap between the lowest-energy undesired state and the desired target state (Fig. 1a). A specificity sweep begins by using ILP to find the sequence with the highest binding affinity for the target, ignoring specificity. An initial value for the quantity Δ is then computed by predicting the energies of all possible complexes involving this design. The ILP optimization is repeated, this time designing a protein that optimizes binding with the target subject to the constraint that all undesired states have energy gaps to the designed state that are larger than Δ plus a small increment. This is repeated, gradually increasing the value of Δ , until it is no longer possible to find design sequences that satisfy the constraints. Although CLASSY can be run with any value assigned to Δ , one advantage of the specificity sweep exploring a broad range of Δ values is that no assumption of how much stability or specificity is ‘enough’ need be made before the calculation.

Candidate designs from a specificity sweep list may be selected for testing by a user, after considering predicted stability/specificity trade-offs and the sequence changes that bring these about. Other considerations may be included, because CLASSY provides the ability

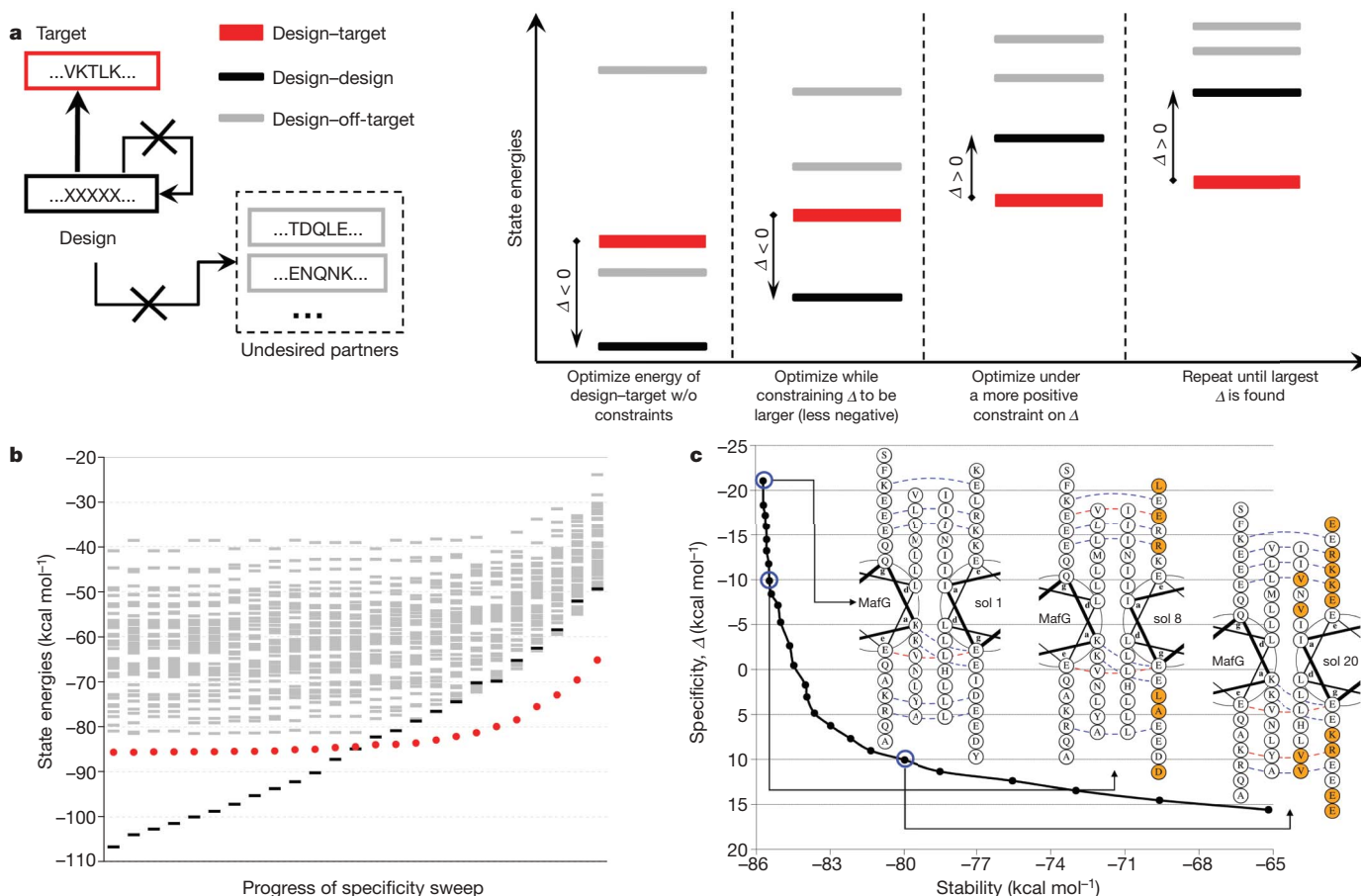


Figure 1 | Designing specific peptides using CLASSY. **a**, Specificity sweep scheme. A sequence (black) is sought that binds a target (red) but not several undesired partners (grey) or itself. Panels from left to right illustrate iterations of the CLASSY procedure, during which the specificity gap Δ is increased. **b**, **c**, A specificity sweep with MafG (also called MAFG) as the target and all other human bZIP coiled coils (except MafK, also called MAFK, which is in the same family as MafG) and the design homodimer as undesired complexes. The plot in **b** corresponds to the cartoon in **a**. Red

dots, black bars and grey bars represent energies of the design–target, design–design and design–other bZIP complexes, respectively. The plot in **c** shows design–target complex stability versus specificity (Δ). Portions of several designed complexes are shown using helical wheels. From left to right these are: the most stable design (sol 1), the eighth solution obtained in the specificity sweep (sol 8) and the twentieth solution obtained (sol 20), which is anti-SMAF. Orange colour highlights amino-acid changes from the previously shown sequence.

to restrict arbitrary linear functions of sequence. In our application, a bias for the bZIP coiled-coil fold was imposed by constraining designs to be leucine-zipper-like according to a position-specific scoring matrix (PSSM). Similar constraints could also be used, for example, to place requirements on predicted solubility. Such considerations, which are often included in designs in an *ad hoc* manner or by using manual post-evaluation and filtering, can be naturally incorporated into the CLASSY procedure.

Design of anti-bZIP peptides

We applied CLASSY to design partners for nearly all human bZIPs and used our computational results to assess the difficulty of the bZIP interaction specificity design problem. We sought anti-bZIP designs predicted to bind their targets and yet interact minimally with themselves and with members of the 19 non-target bZIP families. Because of the extremely high sequence similarity within families, we did not require that the designs discriminate between siblings in the target family. The desired design–target heteromeric complex, as well as undesired design–design and design–off-target complexes, were modelled as coiled-coil dimers on a fixed-backbone template and evaluated using energy functions similar to that of ref. 26, which was shown previously to give good performance predicting native bZIP interaction preferences²⁶ (also see Supplementary Information).

Specificity sweeps were computed for the 46 bZIPs in ref. 4. These calculations predicted that specificity will arise only rarely among bZIP partners optimized for stability alone. Such designs are almost all predicted to form strong homodimers, regardless of the family they are targeted against (Supplementary Fig. 2). Negative design is also required to disfavour complexes with undesired bZIP competitors. Approximately 65% of 46 designs optimized for affinity alone were judged to face significant competition from non-target families; this can be addressed in CLASSY by sacrificing stability, as shown in Supplementary Fig. 2. We carried out additional computational analyses to estimate how candidate bZIP partners are distributed in stability-specificity space (Supplementary Fig. 12). Even when the design–design homodimer is the only undesired state, most sequence space is predicted to be non-specific. Thus, addressing specificity is critical, but the notable reduction this imposes on acceptable sequences makes the design problem challenging.

Testing of anti-bZIP designs

We tested 48 peptides designed to bind representative targets from all 20 bZIP families using a protein microarray assay that has been validated for measuring interaction preferences for bZIPs⁴. Sequences to be tested were selected from the specificity sweeps by hand, considering the magnitude of Δ , the amount of stability lost relative to the most stable design, and sequence features such as excessive loss of hydrophobic interactions in the core. Designs were named after the family of the target they were designed to bind, as defined in Supplementary Fig. 1. See Fig. 1c for the example of anti-SMAF. Supplementary Table 1 provides detailed descriptions of the origin of each design. In a few of the cases where we designed more than one peptide against a given target, experimental results for initial designs were incorporated to guide the CLASSY design procedure. For example, anti-ZF (designed to bind target ZF, which is also called CREBZF) was designed using a modified specificity sweep that up-weighted the influence of XBP-1 (XBP1) in determining Δ , after this protein was experimentally determined to be a problematic competitor. The ability to easily incorporate information about known competitors is one advantage of CLASSY.

In total, 48 peptides designed against 20 targets were tested for interaction with 33 representative human bZIP coiled coils and for self-association. Fluorescence intensities measured on bZIP arrays have previously been shown to reflect relative interaction strengths measured in solution⁴. Each peptide in turn (both designed and native) was labelled with the fluorescent dye Cy-3 and used to probe aldehyde-derivatized slides printed with potential partners. Of the 48 designs

tested, 40 bound to their intended target, as assessed by fluorescence signal above background (Supplementary Fig. 1). The probability of this occurring by chance, given the distribution of signals from designed peptides interacting with human peptides on the arrays, was $\sim 10^{-11}$. Self-association of the designs was also evaluated. Only 40% of the designs showed detectable self-interactions using the same criterion, and all but 6 interacted with a human bZIP more strongly than they interacted with themselves (Fig. 2a and Supplementary Fig. 1).

To determine the interaction specificity of the designed molecules, we used Cy-3-labelled designed peptides and compared the array signal for interaction with the target to that for interaction with non-target competitors. Results for the most specific design identified for each of the 20 families, named according to the target family, are shown in Fig. 2a. For ten designs, the strongest interaction observed was with the intended target. Notably, eight of these designs bound their targets with array signals distinctly greater than for any other non-target-family partner (designs anti-ZF, anti-FOS, anti-SMAF, anti-ATF2, anti-JUN, anti-LMAF, anti-XBP1 and anti-ATF4, leftmost in the specificity panel of Fig. 2a). This indicates measurable interaction specificity on the arrays. For two more designs, fluorescence signal for interaction with the target was only marginally greater than that for interaction with one to two other proteins (designs anti-ATF3 and anti-C/EBP γ). Nine other designs bound their targets, but less strongly than they bound to members of other families. For target family PAR, the designed peptide did not show detectable binding above background.

To assess the stability of each design–target interaction, we labelled each native bZIP target with Cy-3 and probed an array containing 33 representative human bZIP peptides as well as the anti-target design. This experiment assayed design–target stability relative to interactions of the target with its native partner(s). The strongest signal was often from the design–target complex, indicating that many designs can be expected to out-compete native partners of the targets, using modest concentrations (summarized in Fig. 2a; complete data in Supplementary Information). Less stable designs can probably be improved through generic strategies such as the addition of acidic extensions, as for the A-ZIPs²⁰.

To validate the array assay, 28 mixtures involving the 7 best designs were characterized in solution using thermal denaturation monitored by circular dichroism. Each designed peptide was tested for interaction with four candidate partners: the intended target of the design, the next-best interaction partner as reported by the array, a protein closely related by sequence to the target, and itself. We monitored whether the mixtures showed an increase in the temperature of denaturation (T_m) compared to that expected from the average of the signals of the individual components (Fig. 2b–e and Supplementary Figs 3–8). In all cases, the T_m studies supported binding of each design to its intended target. For the 21 undesired complexes tested, 18 either showed no evidence for interaction or a T_m that was clearly lower than that of the design–target complex. For the remaining three undesired complexes, formation of mixtures complicated the analyses, although these are probably also weaker than the corresponding design–target complexes (Supplementary Figs 4–6). Solution data were also examined for consistency with the array measurements and supported the same relative ordering of stabilities for 35 out of 41 comparable cases (see Supplementary Information).

Three of our best designs target c-Jun, c-Fos and ATF-2 (ATF2). These proteins are constituents of the AP-1 transcription factor complexes involved in cell proliferation and oncogenesis. The c-Jun–c-Jun, c-Jun–c-Fos and c-Jun–ATF-2 dimers are involved in these important processes in ways that have not been fully elucidated. Complexes involving c-Jun have previously been targeted for disruption using a dominant-negative A-ZIP version of c-Fos²⁰. However, because c-Fos also binds ATF-2 and its family members⁴, the A-ZIP strategy is not as specific as might be desired. The same is true for c-Jun and ATF-2: native partners of these targets also bind to additional

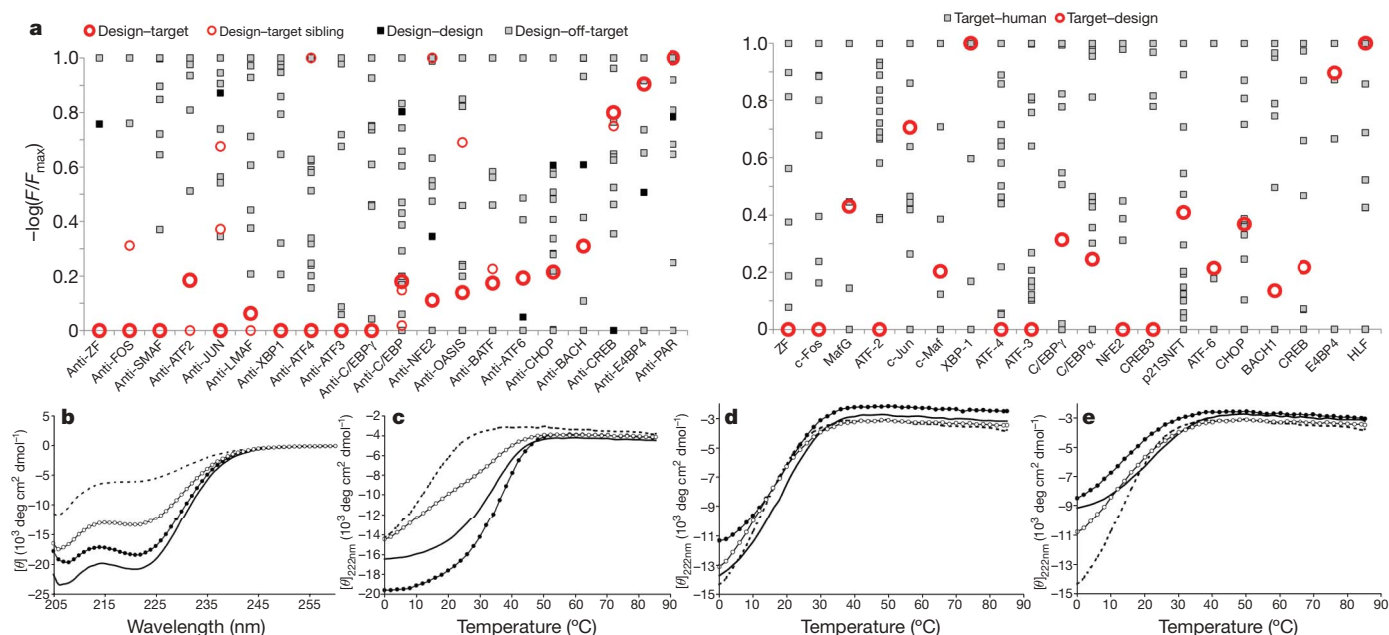


Figure 2 | Experimental testing of anti-bZIP designs. **a**, Peptide array results for the most specific design identified for each human bZIP family. Columns show experiments using the indicated protein to probe an array. For the specificity panel (left), designs in solution were used to probe human bZIPs and designs on the surface. In the relative stability panel (right), human bZIP targets were used to probe an array containing the cognate design of each target and 33 human bZIPs. Data are plotted as $-\log(F/F_{\max})$, where F is the fluorescence signal on the array, such that the strongest interaction has a value of zero. Values of $-\log(F/F_{\max})$ above 1.0 were set to 1.0. Thick red circles, design–target; thin red circles, design interactions with siblings in the target family; grey squares, interactions with other human bZIPs; black squares, design–design. Designs are named using the family of their target, as defined in Supplementary Fig. 1. C/EBP γ (also known as

CEBP γ), C/EBP α (CEBPA), p21SNFT (BATF3), CHOP (DDIT3), CREB (CREB1), E4BP4 (NFIL3). **b–e**, Solution testing of anti-SMAF complexes assayed using circular dichroism. In each panel, anti-SMAF alone is shown with a dashed line, the partner being tested with a solid line, the numerical average of these two signals with open circles and the mixture of the two peptides with filled circles. **b**, **c**, Anti-SMAF interacts with target MafG (T_m of $\sim 38^\circ\text{C}$). **d**, Anti-SMAF interacts, at most, very weakly with c-Jun, the closest competitor according to microarray data. **e**, There is no evidence for anti-SMAF interacting with MafB (MAFB), a sequence closely related to the target. Circular dichroism spectra in **b** were collected at 25°C . Anti-SMAF unfolds with a T_m of $\sim 12^\circ\text{C}$. Similar data for other complexes are included in Supplementary Figs 3–8.

families. Attempts to identify new partners for c-Fos and c-Jun using experimental selection strategies gave peptides that strongly self-associated and also bound bZIPs non-specifically (that is, the intended anti-c-Fos and anti-c-Jun peptides bound both FOS and JUN family members tightly)^{18,30}. Our designed peptides provide a way to introduce specificity, for example, to disrupt c-Jun–c-Fos but not c-Jun–c-Jun or c-Jun–ATF-2, using anti-FOS.

Properties of the anti-bZIP designs

Figure 3a shows the interaction profiles of native bZIP leucine zippers and the designed anti-bZIP peptides. The native proteins exhibit diverse interaction properties, despite their limited sequence variability (Fig. 3b)⁴. The designed peptides are even more limited in sequence diversity, yet they encode many additional specificity profiles, indicating that bZIP-like coiled-coil interaction space is only sparsely sampled by the human proteins (Fig. 3c). The frequency of success of our interaction prediction model, along with results from CLASSY analysis, allowed us to conservatively estimate that $>1,900$ very distinct interaction profiles can be encoded within the restricted sequence space used in our designs. This may prove useful for applications in synthetic biology (see Supplementary Information).

CLASSY designs exhibited canonical bZIP specificity determinants, such as a preference for Asn residues at a positions to pair across helices, and charge complementarity at g -to- e' pairs (see Fig. 1c for coiled-coil heptad positions; a prime indicates a residue on the opposite helix, see Supplementary Fig. 15)^{19,24}. Interestingly, g -to- a' pairs were predicted to make a comparable, if not larger, contribution to specificity than g -to- e' pairs. Other unanticipated specificity patterns also emerged, involving steric interactions between a and d' sites (see Supplementary Information for a fuller discussion). The significance of such

interactions has not been broadly recognized in parallel coiled coils, although recent studies indicate their importance in anti-parallel dimers³¹.

Discussion

CLASSY provides a way to analyse and optimize stability/specificity trade-offs in protein design. The cluster expansion/ILP procedure imposes few formal requirements on the type of scoring function that can be used or the type of specificity problem that can be addressed. However, measuring and predicting interaction specificity for proteins generally remains challenging. Here, the bZIPs provided several advantages. The bZIP microarray assay benefits from reversible folding of short coiled coils, and data from previous array measurements of many bZIP transcription factor pairs were critical for developing predictive models^{4,25,26}. Experimental helix propensities contributed to the quality of these models, and knowledge of particular specificity determinants (for example, the special role of Asn pairs) improved predictions and also disfavoured the formation of higher-order oligomers¹⁹. Finally, symmetric fixed-backbone models proved adequate for this application²⁶. This facilitated both structural modelling and cluster-expansion training, although cluster expansion can also be used for asymmetric structures and with flexible backbones³². Further details about features specific to bZIP modelling are in Methods and Supplementary Discussion.

Determinants of protein-interaction specificity are not yet as well understood for other complexes, but significant progress in this area is evident. Zinc-finger–DNA, SH2–peptide and PDZ–peptide complexes have been extensively studied, and both assays and interaction models have been developed that make these good candidates for design using CLASSY (see Supplementary Information for further

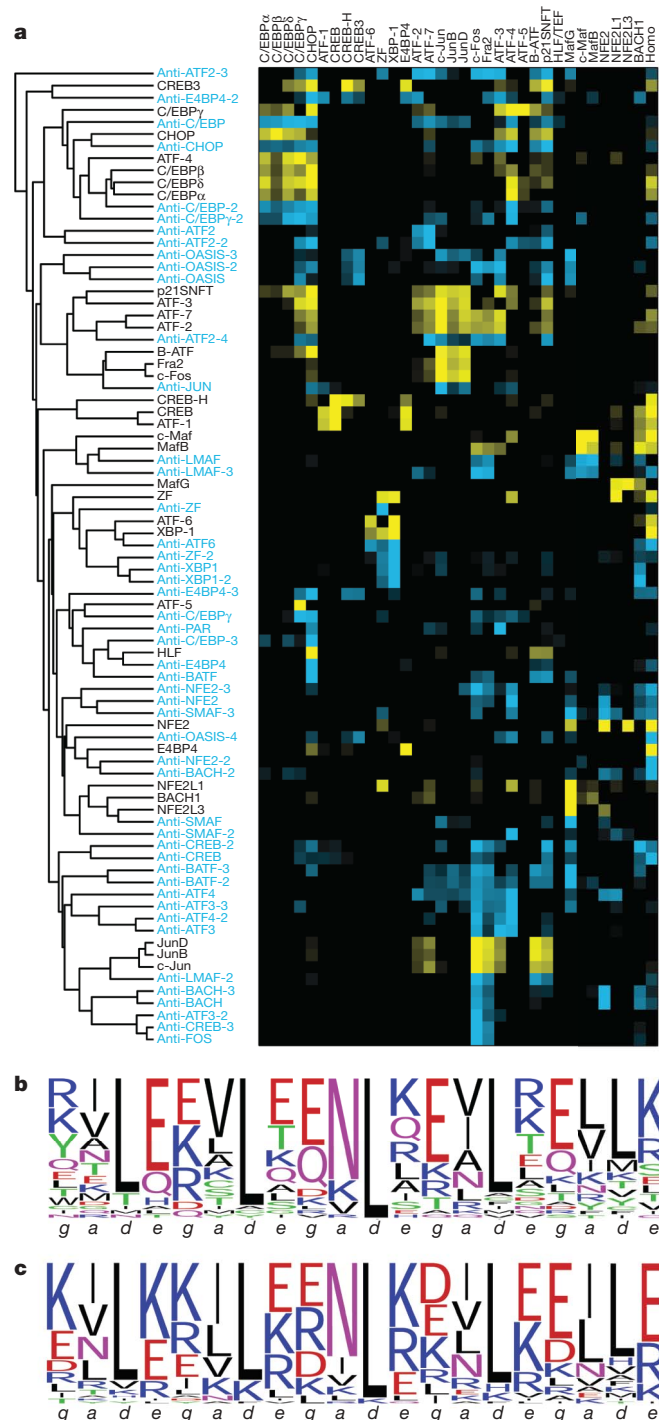


Figure 3 | Properties of designed peptides compared to human bZIP leucine zippers. **a**, Hierarchical clustering of interaction profiles for 33 human peptides and 48 designs; an interaction profile consists of the array signals for interactions with 33 surface-bound human peptides. Proteins on the surface are in columns and those in solution are in rows, with designed proteins and their interaction profiles in blue and human bZIP interaction profiles in yellow. The column labelled Homo at far right indicates self-association of the protein. Alternative names not previously defined are: C/EBP β (CEBPP), C/EBP δ (CEBPD), ATF-7 (ATF7), B-ATF (BATF), Fra2 (FOSL2), CREB-H (CREB3L3), ATF-1 (ATF1), ATF-5 (ATF5), JunD (JUND), JunB (JUNB). **b**, **c**, Sequence logos for *a*, *d*, *e* and *g* positions from the first 5 heptads of the 33 human bZIP leucine zippers in **b** and the 48 designed peptides in **c** (<http://weblogo.berkeley.edu>).

discussion)^{2,3,12,33,34}. Large-scale interaction experiments are becoming more common, and general-purpose models to describe protein structures and energies are under development^{33,35–37}. Advances in

these areas will expand the problems that can be addressed using CLASSY. In the long term, we hope this approach will help to address how interaction crosstalk can be controlled in both evolved and designed protein systems.

METHODS SUMMARY

Structure-based modelling of coiled-coil interactions was done as described previously, with modifications detailed in the Methods and Supplementary Information²⁶. Using the technique of cluster expansion, structure-based models were converted to functions of sequence that included constant, single-residue and residue-pair terms. Training of the cluster expansion used 61,780 random bZIP-like sequences that were modelled structurally^{28,29}. A limited amino-acid alphabet was considered, which included the ten residues most frequently found at each coiled-coil heptad position in native bZIPs. Constrained optimization using ILP was used to design *a*, *d*, *e* and *g* sites. ILP optimization minimized the energy of design–target complexes, subject to constraints on the energy gap with respect to undesired complexes and the match of the design sequence to a position-specific scoring matrix derived from 432 native bZIP leucine zippers. Other positions in the coiled-coil repeat (*b*, *c* and *f* positions) were chosen to be consistent with the designed interface *a*, *d*, *e* and *g* residues, using a probabilistic framework. For each design target, the ILP optimization was repeated with increasing values of the specificity gap parameter Δ , in a procedure termed a specificity sweep. Sequences for experimental testing were selected manually from candidates generated using the specificity sweeps.

For experimental testing, His₆-tagged peptides were expressed in RP3098 cells and purified by Ni-NTA followed by reverse-phase HPLC. Coiled-coil microarrays were printed, processed and probed as described previously⁴. Fluorescence signals from the arrays were processed to remove background and normalized. Circular dichroism measurements were performed using standard techniques to measure spectra between 195 nm and 280 nm at 25 °C or thermal stability by monitoring ellipticity at 222 nm. Data were fit to appropriate thermodynamic equations to obtain apparent T_m values. Detailed descriptions of all procedures are included in Methods and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 August 2008; accepted 9 February 2009.

1. Stiffler, M. A. *et al.* PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **317**, 364–369 (2007).
2. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
3. Wiedemann, U. *et al.* Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol.* **343**, 703–718 (2004).
4. Newman, J. R. & Keating, A. E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **300**, 2097–2101 (2003).
5. Landgraf, C. *et al.* Protein interaction networks by proteome peptide scanning. *PLoS Biol.* **2**, E14 (2004).
6. Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G. & Laub, M. T. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol.* **3**, e334 (2005).
7. Havranek, J. J. & Harbury, P. B. Automated design of specificity in molecular recognition. *Nature Struct. Biol.* **10**, 45–52 (2003).
8. Kortemme, T. *et al.* Computational redesign of protein–protein interaction specificity. *Nature Struct. Mol. Biol.* **11**, 371–379 (2004).
9. Ali, M. H. *et al.* Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed α/β structure. *Structure* **13**, 225–234 (2005).
10. van der Sloot, A. M. *et al.* Designed tumor necrosis factor-related apoptosis-inducing ligand variants initiating apoptosis exclusively via the DR5 receptor. *Proc. Natl Acad. Sci. USA* **103**, 8634–8639 (2006).
11. Yin, H. *et al.* Computational design of peptides that target transmembrane helices. *Science* **315**, 1817–1822 (2007).
12. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nature Struct. Biol.* **9**, 621–627 (2002).
13. Shifman, J. M. & Mayo, S. L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc. Natl Acad. Sci. USA* **100**, 13274–13279 (2003).
14. Fu, X., Appgar, J. R. & Keating, A. E. Modeling backbone flexibility to achieve sequence diversity: the design of novel α -helical ligands for Bcl-xL. *J. Mol. Biol.* **371**, 1099–1117 (2007).
15. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. Specificity versus stability in computational protein design. *Proc. Natl Acad. Sci. USA* **102**, 12724–12729 (2005).
16. Kangas, E. & Tidor, B. Electrostatic specificity in molecular ligand design. *J. Comput. Phys.* **112**, 9120–9131 (2000).

17. Deutsch, J. M. & Kurosky, T. New algorithm for protein design. *Phys. Rev. Lett.* **76**, 323–326 (1996).
18. Mason, J. M., Schmitz, M. A., Muller, K. M. & Arndt, K. M. Semirational design of Jun-Fos coiled coils with increased affinity: universal implications for leucine zipper prediction and design. *Proc. Natl Acad. Sci. USA* **103**, 8989–8994 (2006).
19. Vinson, C., Acharya, A. & Taparowsky, E. J. Deciphering B-ZIP transcription factor interactions *in vitro* and *in vivo*. *Biochim. Biophys. Acta* **1759**, 4–12 (2006).
20. Gerdes, M. J. *et al.* Activator protein-1 activity regulates epithelial tumor cell identity. *Cancer Res.* **66**, 7578–7588 (2006).
21. Krylov, D., Olive, M. & Vinson, C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J.* **14**, 5329–5337 (1995).
22. Acharya, A., Rishi, V. & Vinson, C. Stability of 100 homo and heterotypic coiled-coil a–a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* **45**, 11324–11332 (2006).
23. Krylov, D., Barchi, J. & Vinson, C. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* **279**, 959–972 (1998).
24. Lupas, A. N. & Gruber, M. The structure of α -helical coiled coils. *Adv. Protein Chem.* **70**, 37–78 (2005).
25. Fong, J. H., Keating, A. E. & Singh, M. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.* **5**, R11 (2004).
26. Grigoryan, G. & Keating, A. E. Structure-based prediction of bZIP partnering specificity. *J. Mol. Biol.* **355**, 1125–1142 (2006).
27. Kingsford, C. L., Chazelle, B. & Singh, M. Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics* **21**, 1028–1036 (2005).
28. Grigoryan, G. *et al.* Ultra-fast evaluation of protein energies directly from sequence. *PLOS Comput. Biol.* **2**, e63 (2006).
29. Zhou, F. *et al.* Coarse-graining protein energetics in sequence variables. *Phys. Rev. Lett.* **95**, 148103 (2005).
30. Mason, J. M., Muller, K. M. & Arndt, K. M. Positive aspects of negative design: simultaneous selection of specificity and interaction stability. *Biochemistry* **46**, 4804–4814 (2007).
31. Hadley, E. B., Testa, O. D., Woolfson, D. N. & Gellman, S. H. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proc. Natl Acad. Sci. USA* **105**, 530–535 (2008).
32. Apgar, J. R., Hahn, S., Grigoryan, G. & Keating, A. E. Cluster-expansion models flexible-backbone protein energetics. *J. Comput. Chem.* (in the press).
33. Sanchez, I. E. *et al.* Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. *PLOS Comput. Biol.* **4**, e1000052 (2008).
34. Kaplan, T., Friedman, N. & Margalit, H. Ab initio prediction of transcription factor targets using structural knowledge. *PLOS Comput. Biol.* **1**, e1 (2005).
35. Boas, F. E. & Harbury, P. B. Potential energy functions for protein design. *Curr. Opin. Struct. Biol.* **17**, 199–204 (2007).
36. Das, R. & Baker, D. Macromolecular modeling with ROSETTA. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
37. Zhou, H. & Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* **11**, 2714–2726 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NIH award GM67681 and used computer equipment purchased under the NSF award 0216437. We thank the MIT BioMicro center for arraying instrumentation and R. T. Sauer, M. Singh, B. Tidor, M. Laub, T. A. Baker, J. H. Davis, M. S. Kay, J. R. S. Newman, W. F. DeGrado and members of the Keating laboratory, especially O. Ashenberg and T. C. S. Chen, for comments on the manuscript.

Author Contributions G.G., A.W.R. and A.E.K. conceived the project. G.G. developed, implemented and applied the CLASSY formalism and carried out all computational analyses. A.W.R. designed and performed all experiments. All authors analysed data and guided the research plan. G.G. and A.E.K. wrote the paper, in consultation with A.W.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.E.K. (keating@mit.edu).

METHODS

Modelling bZIP leucine zipper interactions. Two variants of the previously described energy function HP/S/C were used to evaluate the relative stability of coiled-coil dimer structures²⁶. Models were constructed using a single backbone, with rotameric sampling and continuous relaxation used to position side chains. The first energy function, HP/S/Ca, is as published²⁶, with scale factor $s = 0$ such that intra-chain interactions in the folded structure do not directly contribute to stability (although there are indirect contributions). HP/S/Ca replaces core $a-a'$ and $d-d'$ terms derived from structure-based calculations with weights from a machine-learning algorithm²⁶. In the variant model HP/S/Cv, structure-based $a-a'$ interactions were replaced with $a-a'$ experimental coupling energies for 55 amino-acid combinations²² and the $d-d'$ interaction for Leu–Leu was replaced with the empirical value -2 kcal mol^{-1} . Following cluster expansion (see below), a -position point contributions were adjusted such that 100 folding free energies measured previously²² were predicted optimally (in the least-squares sense, see Supplementary Fig. 10). The following ten amino acids were allowed: V, L, N, I, K, A, R, T, S and E for a positions; L, V, I, M, H, Y, T, A, K and F for d positions; E, K, R, Q, L, S, T, A, V and I for e positions; E, K, Q, R, L, Y, T, D, A and I for g positions. These are the ten amino acids most frequently encountered in the respective positions in bZIPs. Additionally, for the a position, these are also the ten amino acids for which coupling energies have been measured previously²².

Cluster expansion. Cluster expansion provides a way to express the energy of a sequence adopting a particular backbone structure as an algebraic function of the sequence itself²⁸. The formal basis of the technique is described in Supplementary Methods. In this study, the desired and undesired structures had the same backbone, and thus one cluster expansion (for parallel, coiled-coil dimers) was sufficient. Cluster expansion calculations for both HP/S/Ca and HP/S/Cv included single-residue and residue-pair terms. A training set was built by randomly generating 61,780 coiled-coil sequences with heptad position-specific amino-acid probabilities taken from a multi-species alignment of 432 bZIPs (M. Singh, personal communication). Gly and Pro were not included. Pair contributions were included only for amino-acid pairs ≤ 7 residues apart, resulting in 9,929 possible effective cluster interactions: 1 constant, 68 point and 9,860 pair terms. After the fitting procedure, 2,544 and 2,470 effective cluster interactions survived the statistical significance test (that is, lowered the cross-validated error²⁸) for HP/S/Ca and HP/S/Cv, respectively. The performance of the resulting cluster expansions on a similarly generated test set of 10,000 sequences not used in training is shown in Supplementary Fig. 11.

Multi-state design optimization. Design sequences were optimized for interaction with the target using ILP, imposing constraints on the design interaction energy with competitors and a degree of match to a bZIP PSSM. The ILP and PSSM are detailed in Supplementary Methods. We performed two types of CLASSY calculations. The first, a specificity sweep, starts by using ILP to identify the design sequence that produces the provably lowest predicted binding energy to a target. Given this sequence, energy gaps between the design–target dimer and all design–competitor dimers, including design–design, are calculated as $\text{gap} = E_{\text{design-competitor}} - E_{\text{design-target}}$. The minimal gap (which may be negative) is defined as Δ . In the next iteration of the specificity sweep, the design–target energy is re-optimized, this time imposing constraints to require that all gaps be greater than $\Delta + 1 \text{ kcal mol}^{-1}$. In each round, Δ is updated and this procedure is repeated until no more solutions exist (Fig. 1a). Designs to be tested are chosen from this list of optimized sequences, as discussed in the main text.

Anti-bZIP designs were tested in three rounds of microarray experiments. When we sought to improve upon a previously tested design, we sometimes used experimental results to formulate biased specificity sweeps. In these calculations, custom offsets were applied to enhance or diminish the significance of some gaps relative to others; the remainder of the procedure was identical to that for a standard specificity sweep. For example, a biased specificity sweep was used to design anti-ZF after the first design tested (named as anti-ZF-2) interacted with XBP-1 more strongly than with ZF, contrary to predictions of the model. This is illustrated and explained further in Supplementary Fig. 9. Supplementary Table 1 contains a list of all designs and the procedures by which they were obtained, including the details of any biased specificity sweeps used.

In all CLASSY procedures, except where noted in Supplementary Table 1, 46 human bZIPs were considered (sequences taken from ref. 25), and the modelled states were as follows: the design–target complex was the only desired state; design–off-target bZIP complexes for all bZIPs not in the family of the target bZIP were treated as undesired; the design–design homodimer was also an undesired state.

Further details regarding the theory behind CLASSY, as well as other computational analyses performed in this study, are in Supplementary Methods and Supplementary Discussion.

Choosing 33 representative human bZIPs. To avoid redundancy and to conserve resources and time, we used a representative set of 33 human bZIPs that covered all 20 families (see Supplementary Fig. 13). Representatives were chosen based on sequence similarity and reagent availability and described well the distinct interaction profiles reported previously⁴. Computational design was nevertheless conducted with 46 human bZIPs taken from ref. 4.

Plasmid construction and peptide expression, purification and labelling. Synthetic genes encoding all designs were constructed using DNAWorks³⁸ to design primers that contained flanking BamHI and XhoI restriction sites. A two-step PCR method was used to assemble the primers, and the PCR products were digested with BamHI and XhoI and cloned into a modified pDEST17 vector⁴. All synthetic genes were confirmed to be correct by sequencing. Plasmids encoding human leucine-zipper peptides have been published previously⁴, with the exception of modified sequences for JUN family constructs that are described in the Supplementary Methods.

Plasmids were transformed into RP3098 cells and 1-l cultures in LB were grown to 0.4–0.6 OD and induced at 37 °C for 3–4 h with addition of 1 mM IPTG. Peptides were purified under denaturing conditions (guanidine hydrochloride, GuHCl) by binding to Ni-NTA resin and eluted with 60% acetonitrile/1% TFA. After reduction with 10 mM TCEP in 5% acetic acid for 3 min at 65 °C, peptides were further purified using reverse-phase HPLC. The molecular weights of all designed peptides were confirmed as correct to within 0.15% by mass spectrometry. To generate dye-labelled peptides, 10 molar excess of Cy3 NHS ester in 6 M GuHCl/100 mM phosphate (pH 7.5) was added to lyophilized aliquots of protein and incubated for 2 h at 22–25 °C. Free dye was removed using size-exclusion spin columns. Labelled peptides were stored at -80°C .

Preparation and probing of arrays. Lyophilized aliquots of protein were resuspended to a concentration of 40 μM in 6 M GuHCl, 100 mM phosphate (pH 7.5), 0.04% Tween-20, 10 μM Alexa Fluor 633 hydrazide. Proteins were printed on aldehyde-presenting glass slides (Thermo Fisher Scientific) using a Microgrid TAS Arrayer. Twelve identical subarrays were printed on each slide. Each protein was spotted twice, in two different print orders, for a total of four spots for each protein per subarray. After printing, slides were divided into subarrays by drawing a hydrophobic boundary (PAP pen, Electron Microscopy Sciences). Slides were stored at -80°C for up to 1 month.

Slides were prepared for probing by: washing face up in -80°C ethanol for 30 s; transferring to 80% ethanol/10 mM NaOH and incubating with shaking for 15 min; washing in H_2O for 15 s; incubating in PBS/0.1% Tween-20 for 15 min with shaking; and drying by centrifugation. Slides were then immediately probed by diluting labelled peptide in 6 M GuHCl/100 mM phosphate (pH 7.5)/6 mM TCEP sixfold into 1.2 \times buffer (1.2% BSA, 1.2 \times PBS, 0.12% Tween-20). The resulting solution was mixed and 35 μl was immediately pipetted onto each subarray. Each sample was probed in duplicate on adjacent subarrays, for a total of 8 spots used to detect each interaction. Slides were covered with a box and incubated for 1 h. Slides were washed in PBS/0.1% Tween-20 for 15 s and then H_2O for 15 s, and were then dried by centrifugation. Slides were scanned using a DNA Microarray Scanner (Agilent) at several photo-multiplier tube voltage levels. The concentration of probe was 160 nM unless otherwise indicated.

Additional details on experimental techniques and data analysis are provided in Supplementary Methods.

38. Hoover, D. M. & Lubkowsky, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* 30, e43 (2002).

A massive hypergiant star as the progenitor of the supernova SN 2005gl

A. Gal-Yam¹ & D. C. Leonard²

Our understanding of the evolution of massive stars before their final explosions as supernovae is incomplete, from both an observational and a theoretical standpoint. A key missing piece in the supernova puzzle is the difficulty of identifying and studying progenitor stars. In only a single case—that of supernova SN 1987A in the Large Magellanic Cloud—has a star been detected at the supernova location before the explosion, and been subsequently shown to have vanished after the supernova event¹. The progenitor of SN 1987A was a blue supergiant, which required a rethink of stellar evolution models². The progenitor of supernova SN 2005gl was proposed to be an extremely luminous object³, but the association was not robustly established (it was not even clear that the putative progenitor was a single luminous star). Here we report that the previously proposed³ object was indeed the progenitor star of SN 2005gl. This very massive star was likely a luminous blue variable that standard stellar evolution predicts should not have exploded in that state.

On 2007 September 26.7, we used the Hubble Space Telescope (HST) to observe the location of SN 2005gl within its host galaxy NGC 266, in order to confirm or reject that a luminous point source previously identified in pre-explosion images from 1997³ as being spatially coincident with this supernova was in fact the progenitor star. In the new images, the point source NGC266_LBV 1 is no longer visible (Fig. 1), confirming that it was, indeed, the progenitor of SN 2005gl. Before our observation, the proposal that NGC266_LBV 1 was the progenitor star of SN 2005gl was subject to two major caveats. First, earlier HST images could not rule out the possibility that the point source detected in 1997 was a compact cluster of many stars, with a combined luminosity of the order of 10^6 solar luminosities ($10^6 L_{\odot}$). Second, the spatial coincidence of the putative single star with SN 2005gl alone did not provide conclusive evidence that the supernova explosion was actually related to the luminous star—the explosion could have been the result of the death of a lower-luminosity star, projected close to NGC266_LBV 1, but undetectable in the 1997 HST images. Our new observations from 2007 (Fig. 1c), which are substantially deeper than those obtained in 1997, show no trace of SN 2005gl, nor of NGC266_LBV 1, to a deep limit: $V > 25.6$ mag at the 3σ confidence level ($V > 25.9$ mag at the 2σ level; Supplementary Information section 2). These new data attend to both caveats noted above, and show that NGC266_LBV 1 was a single star and that it indeed vanished following the explosion of SN 2005gl, as is expected for a supernova progenitor star. On the basis of its luminosity, such a star is likely to be an extreme member of the group of luminous blue variable stars (LBVs), which are thought to be very massive ($> 50 M_{\odot}$, where M_{\odot} is the solar mass) short-lived stars⁴.

Our spectral analysis (Fig. 2) reveals signatures for interaction between the supernova ejecta and a mass shell, previously lost from the progenitor star, which is the defining feature of a supernova of type II_n (refs 5, 6). Thus, our recent data prove that the type II_n

supernova 2005gl resulted from the explosion of a most luminous, and probably very massive, progenitor star. Following SN 1987A, this is only the second case for which such solid evidence exists, with SN 2005gl being more than 1,000 times more distant than SN 1987A (66 Mpc versus 50 kpc, respectively), and NGC266_LBV 1 being approximately ten times more luminous than Sk −69 202, the blue supergiant progenitor of SN 1987A¹.

During the past decades, several objects detected in pre-explosion images of the locations of nearby supernovae have been proposed as the likely progenitor stars^{1,7–11}. In all of these cases, the masses of the proposed progenitors are estimated to be in the range 8–20 M_{\odot} (ref. 11), significantly below the implied mass of the luminous NGC266_LBV 1. Furthermore, with the exception of SN 1987A, only

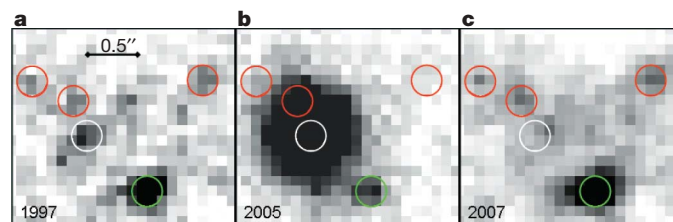


Figure 1 | The progenitor star of SN 2005gl vanished after the supernova event. **a**, A section from the reduced HST image of the galaxy NGC 266, taken on 1997 June 15.1 UT (8 years before the discovery²⁷ of SN 2005gl on 2005 October 5.18 UT), with the Wide Field and Planetary Camera 2 (WFPC2) using the F547M filter (in the green part of the visible light spectrum, centred at 547 nm), produced using Pydrizzle software (Supplementary Information section 1). North is up and east is due left. The luminous star NGC266_LBV 1 (white circle) is detected as a 6.2σ point source, with a flight magnitude of $V = 24.1 \pm 0.1$ mag, which translates to an absolute magnitude $m_V = -10.3$ (a luminosity of $1.1 \times 10^6 L_{\odot}$ at the distance³ of NGC 266 (66 Mpc). Green and red circles are defined below. **b**, A high-resolution image of SN 2005gl obtained using the Near Infrared Camera 2 (NIRC2) behind the laser-guide-star-assisted adaptive optics system at the Keck II telescope in Mauna Kea, Hawaii, on 11 November 2005. The image of the supernova is registered to the pre-explosion HST image of the same area (**a**) using common point sources, to a precision of better than a fraction of a WF pixel³. SN 2005gl is precisely coincident with NGC266_LBV 1, with a formal offset of $0.02''$, well within the positional uncertainties of these objects. **c**, New HST imaging of this location obtained on 26 September 2007 UT, using the same instrument and filter as in 1997 (WFPC2/F547M) with a total exposure time of $4 \times 400 = 1,600$ s, reduced in a similar manner to the data from 1997 and registered onto the 1997 coordinate grid. Green circle shows a bright source near the supernova location easily visible in **a–c**; additional nearby sources (three examples marked in red circles) can be seen in both the new and old visible-light HST data (**a**, **c**) but not in the near-infrared (**b**). As can be seen by comparing **a**, **b** and **c**, SN 2005gl is no longer detected, and NGC266_LBV 1 has also vanished (compare with the nearby sources marked with red circles). These new data provide compelling evidence that the luminous star NGC266_LBV 1 indeed exploded as SN 2005gl in 2005.

¹Benoziyo Center for Astrophysics, Faculty of Physics, The Weizmann Institute of Science, Rehovot 76100, Israel. ²Department of Astronomy, San Diego State University, San Diego, California 92182, USA.

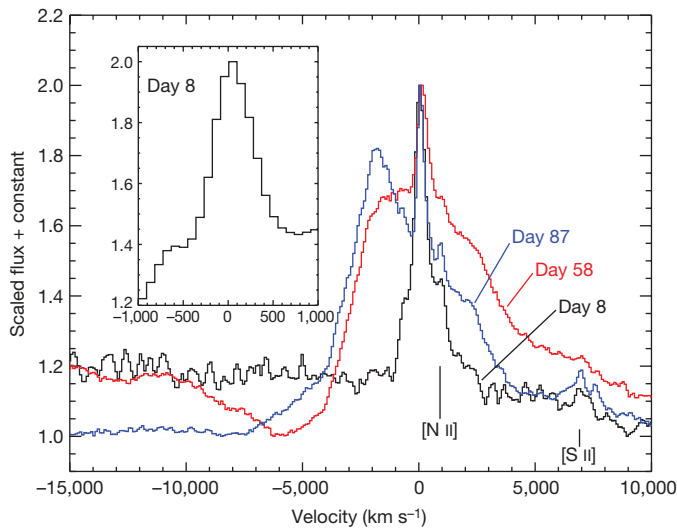


Figure 2 | Spectra of SN 2005gl centred on the region of the H α emission line, presented in velocity scale. The velocity zero point is defined assuming $\lambda_0 = 6,562.85$ Å for this line and a systemic velocity $v = 4,661$ km s $^{-1}$ for NGC 266. Negative velocity values indicate blue-shifted emission coming from material moving towards the observer, while positive values indicate red-shifted emission from material moving away. The spectral resolutions are 3.4 Å for day 8, and 7 Å for days 58 and 87 (155 km s $^{-1}$ and 320 km s $^{-1}$ at H α , respectively). The first-epoch spectrum (day 8; magnified view in inset) shows a narrow, resolved component of H α , with $v \approx 420$ km s $^{-1}$, superposed on an intermediate-width component, with $v \approx 1,500$ km s $^{-1}$. (Note that velocities are determined from the half-width at zero-intensity (HWZI) of the line profiles, and have been corrected for the instrumental resolution of the spectrum at H α (~ 155 km s $^{-1}$).) Narrow nebular lines of nitrogen and sulphur, most probably from unresolved H II region light, are also detected. The narrow component of H α is probably also contaminated by unresolved H II region emission, but its width indicates a contribution from a mass shell ejected shortly before the supernova explosion. The velocity is typical of LBV eruptions, but is well in excess of typical red supergiant wind speeds (Supplementary Information section 5). The expansion velocity of the intermediate-width component probably corresponds to the velocity of the outgoing supernova blast wave ploughing through the slower pre-explosion mass loss, having been decelerated from an initial velocity that is typically ten times larger. No broad component is detected at this epoch, indicating that the photosphere effectively lies at the shock interaction zone, and shields the inner areas (including the fast, unshocked and freely expanding supernova ejecta and the supernova photosphere) from view. From the combination of the line widths and the intermediate-width line luminosity one can calculate the mass loss rate from NGC266_LBV 1 before its explosion, $\dot{M} \approx 0.03 M_{\odot} \text{ yr}^{-1}$ (see main text). The derived mass-loss-rate value is corrected for Galactic extinction in the direction of SN 2005gl of $A_V = 0.23$ mag and an extinction range $A_V < 0.1$ mag in NGC 266. The second spectrum (day 58) has evolved dramatically. A broad component of the H α emission line now dominates, with the intermediate-width component no longer evident. This indicates that the strong interaction phase is probably over, with the emission now dominated by fast, unshocked ejecta ($\sim 10,000$ km s $^{-1}$, a typical speed for a supernova at this age) and the supernova photosphere. The spectrum now resembles that of non-interacting supernovae from lower-mass stars (type II-P events, Supplementary Information section 3; Supplementary Fig. 4). A narrow H α emission remains visible, probably including contributions from both unrelated H II region light, as well as from photoionized unshocked wind surrounding the supernova; at the resolution of our last two spectra the line is only marginally resolved, so it is not possible to accurately determine a velocity for the material and thus discriminate between the two possibilities. The last spectrum (day 87) is nebular, with a wide, asymmetric H α emission profile.

spatial coincidence supports the proposed progenitor identifications; that is, these objects could be, at least in principle, unrelated objects superposed near our line of sight to the supernova by chance^{11–13}.

The bright explosion of a supernova serves as a backlight that illuminates evidence for the last stages of the evolution of the progenitor star before explosion. We now demonstrate this for the

case of SN 2005gl, following the recent¹⁴ general description of the physics of a supernova blast wave expanding into a previously ejected mass shell (see also references within ref. 14). A detailed reanalysis of our spectroscopy of SN 2005gl¹³ (Fig. 2; Supplementary Information) in this context reveals the following. Initially, the spectrum is dominated by a resolved narrow component (Fig. 2 inset) of the Balmer H α emission line, superposed on an intermediate-width component. These probably arise from an unshocked outflowing shell of material ejected by the progenitor shortly before explosion, at a velocity of ~ 420 km s $^{-1}$ (narrow component), and the outgoing supernova blast wave ploughing through the slower pre-explosion mass loss at $\sim 1,500$ km s $^{-1}$ (intermediate component) acting as the effective photosphere. Our next spectrum, taken 50 days later, shows only a broad component of the H α emission line, while the intermediate-width component is gone, indicating that the strong interaction phase is probably over, with the emission now dominated by fast, unshocked ejecta ($\sim 10,000$ km s $^{-1}$). From nebular emission lines of nitrogen and oxygen near the supernova location, we can measure the amount of heavy elements in the gaseous environment of NGC266_LBV 1 (Supplementary Information section 4). Assuming that this is representative of the heavy element content of the short-lived progenitor star itself, we find that the progenitor of SN 2005gl was quite metal-rich, with an oxygen abundance that was $2.8^{+2.7}_{-1.4}$ times that of the Sun.

Assuming that the supernova blast wave went through the surrounding pre-explosion outflow at a velocity of $v_{\text{shock}} \approx 1,500$ km s $^{-1}$ and essentially had emerged from it by day 58, and given the outflow velocity was $v_{\text{wind}} \approx 420$ km s $^{-1}$, the final mass loss episode must have begun no more than about half a year before the final explosion of the star. We use these velocities, and the luminosity of the intermediate-width component of the H α emission line ($L_{\text{int}} = 2.8 \times 10^{39}$ erg s $^{-1}$ cm $^{-2}$; Supplementary Information section 5)—which presumably arises from the shocked gas layer—to derive the mass loss rate \dot{M} during the final, pre-supernova outburst of the progenitor star through the formula¹⁵ $L_{\text{int}} = \epsilon_{\text{H}\alpha} \dot{M} v_{\text{shock}}^3 / 4 v_{\text{wind}}$. Adopting an efficiency factor $\epsilon_{\text{H}\alpha} = 0.1$, appropriate for young supernovae¹⁵, we find a value of $\dot{M} \approx 0.03 M_{\odot} \text{ yr}^{-1}$, which is similar, for example, to that of the LBV P Cygni during its giant eruptions. As the final mass-loss episode lasted for less than a year, our data suggest that the total amount of mass lost just before explosion was modest, and was overrun by the supernova ejecta early on (compare ref. 14), which is in accord with the unremarkable luminosity and moderately rapid decline of the light curve of this supernova³.

The fact that a very luminous and hence very massive star exploded during an LBV-like phase requires a modification in theoretical evolutionary models of extremely massive stars, which generally predict that massive stars should explode after the LBV phase has ended^{16,17}. Furthermore, such models predict that high metallicity stars, similar to that of NGC266_LBV 1, should lose their hydrogen-rich outer envelopes and evolve into stripped hydrogen-poor stars, and end their lives in hydrogen-free supernova explosions some 10^5 years later. Evidently, this was not the fate of NGC266_LBV 1. Our observations thus bolster previous suggestions^{3,18–20} that at least some massive hypergiant stars explode during, or shortly after, a violent LBV-like eruption phase, as hydrogen-rich supernovae in which the ejecta interact almost immediately with the slowly moving shells from prior eruptions.

The lack of an extended plateau in the optical light curve of SN 2005gl¹³ indicates that the progenitor star did not retain a substantial hydrogen envelope, as models suggest that as little as $10 M_{\odot}$ are enough to support a plateau phase²¹. We have shown that NGC266_LBV 1 lost only a modest amount of mass ($< 0.1 M_{\odot}$) in the eruption episode immediately before its final explosion. Therefore, although NGC266_LBV 1 was initially composed mostly of hydrogen, it must have lost many tens of solar masses of hydrogen in previous mass ejection episodes, which may have manifested as a variety of luminous supernova-like events that did not destroy the star^{22,23} (Fig. 3). In any case, our observations strongly constrain the mass loss properties of a very massive star before explosion, which

relevant models must recover. According to standard models²⁴, the likely remnant of the explosion of such a massive star, suffering only limited mass loss (as it retained enough hydrogen to make a type II supernova), is a stellar-mass black hole.

Based on the solid foundation laid by the unambiguous association of the type II_n supernova SN 2005gl with a very luminous hypergiant progenitor undergoing LBV-like mass loss, a synthesis of the entire range of related phenomena (for example, LBV eruptions, supernova

precursors²⁵ and impostors²⁶, putative shell-shell supernova-like events²³ and the final explosions; Fig. 3) can be used to map out the violent lives and deaths of the most massive stars.

Received 12 July 2008; accepted 11 February 2009.

Published online 22 March 2009.

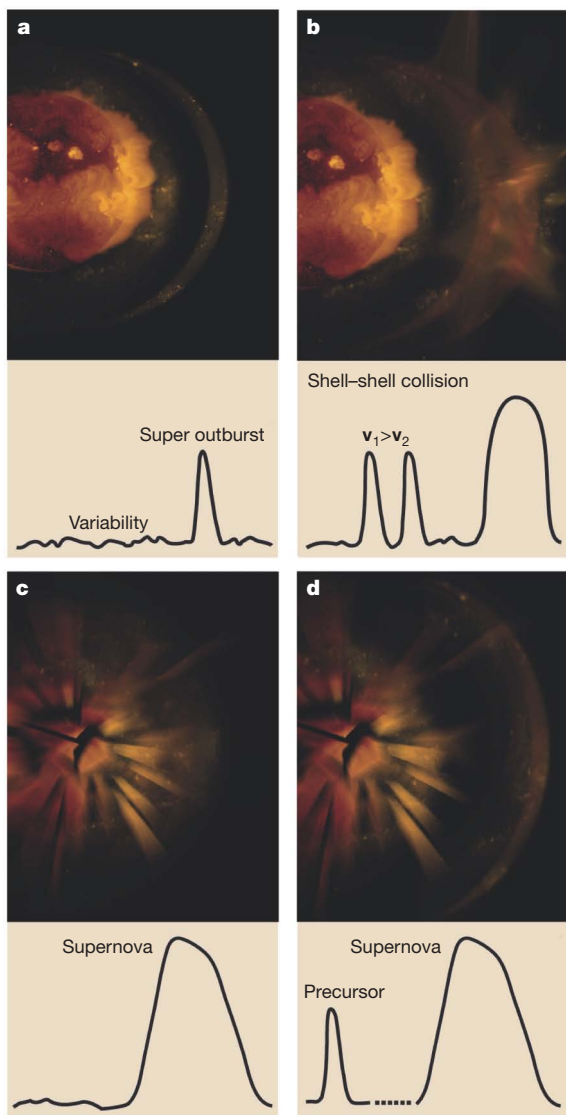


Figure 3 | Graphic illustrations of various manifestations of the violent evolution of the most luminous stars. **a**, LBVs suffer frequent eruptions (with a typical mass-loss rate of around $10^{-2} M_{\odot} \text{ yr}^{-1}$) accompanied by relatively low-level (of the order of 50%) variability⁴ (line graphs at the bottom of panels illustrate temporal flux evolution, with time flowing due right). Less frequent ‘super outbursts’ may involve the ejection of massive, many-solar-mass shells^{4,28}, along with very luminous optical displays, during which the luminosity of the star may increase by an order of magnitude or more. These events (sometimes called supernova impostors²⁶) can therefore be confused with genuine supernova explosions. **b**, Collisions between a faster massive ejected shell (v_1) and a slower one (v_2) have been speculated to result in very luminous events, comparable energetically to bright supernova explosions, but which do not lead to a total destruction of the star^{22,23}. **c**, Our observations show that such luminous stars can evolve and explode as supernovae, after having previously lost some, but not all, of their hydrogen envelopes. The interaction of the expanding debris from the supernova explosion with previously lost mass will result in strong shocks, producing the typical signatures of type II_n supernovae. **d**, Super outbursts occurring shortly before the final explosion of such stars can appear as supernova precursors²⁵.

- Gilmozzi, R. *et al.* The progenitor of SN1987A. *Nature* **328**, 318–320 (1987).
- Arnett, W. D., Bahcall, J. N., Kirshner, R. P. & Woosley, S. E. Supernova 1987A. *Annu. Rev. Astron. Astrophys.* **27**, 629–700 (1989).
- Gal-Yam, A. *et al.* On the progenitor of SN 2005gl and the nature of Type II_n supernovae. *Astrophys. J.* **656**, 372–381 (2007).
- Humphreys, R. M. & Davidson, K. The luminous blue variables: Astrophysical geysers. *Publ. Astron. Soc. Pacif.* **106**, 1025–1051 (1994).
- Schlegel, E. M. A new subclass of Type II supernovae? *Mon. Not. R. Astron. Soc.* **244**, 269–271 (1990).
- Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).
- Aldering, G., Humphreys, R. M. & Richmond, M. S. N. 1993J: The optical properties of its progenitor. *Astron. J.* **107**, 662–672 (1994).
- Smartt, S. J. *et al.* Detection of a red supergiant progenitor star of a Type II-plateau supernova. *Science* **303**, 499–503 (2004).
- Li, W. *et al.* Identification of the red supergiant progenitor of supernova 2005cs: Do the progenitors of Type II-P supernovae have low mass? *Astrophys. J.* **641**, 1060–1070 (2006).
- Hendry, M. A. *et al.* SN 2004A: Another Type II-P supernova with a red supergiant progenitor. *Mon. Not. R. Astron. Soc.* **369**, 1303–1320 (2006).
- Smartt, S. J., Eldridge, J. J., Crockett, R. M. & Maund, J. R. The death of massive stars – I. Observational constraints on the progenitors of type II-P supernovae. Preprint at (<http://arxiv.org/abs/0809.0403v2>) (2008).
- Leonard, D. C. *et al.* An upper mass limit on a red supergiant progenitor for the Type II-plateau supernova SN 2006my. *Publ. Astron. Soc. Pacif.* **120**, 1259–1266 (2008).
- Roelofs, G., Bassa, C., Voss, R. & Nelemans, G. On the detection of the progenitor of the type Ia supernova 2007on. *Mon. Not. R. Astron. Soc.* **391**, 290–296 (2008).
- Smith, N. *et al.* SN 2006tf: Precursor eruptions and the optically thick regime of extremely luminous Type II_n supernovae. *Astrophys. J.* **686**, 467–484 (2008).
- Salamanca, I. *et al.* The circumstellar medium of the peculiar supernova SN1997ab. *Mon. Not. R. Astron. Soc.* **300**, L17–L21 (1998).
- Maeder, A. & Conti, P. S. Massive star populations in nearby galaxies. *Annu. Rev. Astron. Astrophys.* **32**, 227–275 (1994).
- Langer, N. *et al.* Towards an understanding of very massive stars. A new evolutionary scenario relating O stars, LBVs and Wolf-Rayet stars. *Astron. Astrophys.* **290**, 819–833 (1994).
- Smith, N. *et al.* SN 2006gy: Discovery of the most luminous supernova ever recorded, powered by the death of an extremely massive star like η Carinae. *Astrophys. J.* **666**, 1116–1128 (2007).
- Kotak, R. & Vink, J. S. Luminous blue variables as the progenitors of supernovae with quasi-periodic radio modulations. *Astron. Astrophys.* **460**, L5–L8 (2006).
- Trundle, C., Kotak, R., Vink, J. S. & Meikle, W. P. S. SN 2005 gj: Evidence for LBV supernovae progenitors? *Astron. Astrophys.* **483**, L47–L50 (2008).
- Nadyozhin, D. K. Explosion energies, nickel masses and distances of Type II plateau supernovae. *Mon. Not. R. Astron. Soc.* **346**, 97–104 (2003).
- Woosley, S. E., Blinnikov, S. & Heger, A. Pulsational pair instability as an explanation for the most luminous supernovae. *Nature* **450**, 390–392 (2007).
- Dessart, L., Hillier, D. J., Gezari, S., Basa, S. & Matheson, T. SN 1994W: an interacting supernova or two interacting shells? *Mon. Not. R. Astron. Soc.* **394**, 21 (2009).
- Mazzali, P. A. *et al.* The metamorphosis of supernova SN 2008D/XRF 080109: A link between supernovae and GRBs/hypernovae. *Science* **321**, 1185–1188 (2008).
- Pastorello, A. *et al.* A giant outburst two years before the core-collapse of a massive star. *Nature* **447**, 829–832 (2007).
- Van Dyk, S. D. Supernova impostors: LBV outbursts from the most massive stars. *Highlights Astron.* **14**, 205 (2007).
- Puckett, T. & Ceravolo, P. Supernovae 2005gk and 2005gl. *IAU Circ.* **8615** (2005).
- Smith, N. & Owocki, S. P. On the role of continuum-driven eruptions in the evolution of very massive stars and Population III stars. *Astrophys. J.* **645**, L45–L48 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge advice from E. Nakar, R. Sari, E. Baron, S. Smartt, M. Modjaz and P. Mazzali, and E. Ofek also for his help with Fig. 1. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA. A.G.-Y. acknowledges support by the Israeli Science Foundation, an EU Seventh Framework Programme Marie Curie IRG fellowship, the Benozio Center for Astrophysics, a research grant from the Peter and Patricia Gruber Awards, and the William Z. and Eda Bess Novick New Scientists Fund at the Weizmann Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.G.-Y. (avishay.gal-yam@weizmann.ac.il).

LETTERS

Ballistic spin resonance

S. M. Frolov¹, S. Lüscher¹, W. Yu¹, Y. Ren¹, J. A. Folk¹ & W. Wegscheider²

The phenomenon of spin resonance has had far-reaching influence since its discovery 70 years ago¹. Electron spin resonance driven by high-frequency magnetic fields has enhanced our understanding of quantum mechanics, and finds application in fields as diverse as medicine and quantum information². Spin resonance can also be induced by high-frequency electric fields in materials with a spin-orbit interaction; the oscillation of the electrons creates a momentum-dependent effective magnetic field acting on the electron spin^{3–9}. Here we report electron spin resonance due to a spin-orbit interaction that does not require external driving fields. The effect, which we term ballistic spin resonance, is driven by the free motion of electrons that bounce at frequencies of tens of gigahertz in micrometre-scale channels of a two-dimensional electron gas. This is a frequency range that is experimentally challenging to access in spin resonance, and especially difficult on a chip. The resonance is manifest in electrical measurements of pure spin currents¹⁰—we see a strong suppression of spin relaxation length when the oscillating spin-orbit field is in resonance with spin precession in a static magnetic field. These findings illustrate how the spin-orbit interaction can be harnessed for spin manipulation in a spintronic circuit¹¹, and point the way to gate-tunable coherent spin rotations in ballistic nanostructures without external alternating current fields.

An electron moving along a periodic trajectory, $\mathbf{k}(t)$, experiences an oscillating spin-orbit field, $\mathbf{B}^{\text{so}}(\mathbf{k})$. One way to generate such a periodic trajectory is from specular scattering between two parallel boundaries in a conducting channel (Fig. 1c inset). A spin resonance at frequency f_0 can be expected from this ballistic motion when the frequency of a typical bouncing trajectory matches the spin precession frequency in a magnetic field B_0 :

$$f_0 \equiv \frac{g\mu_B B_0}{h} \approx \frac{v_F}{2w} \quad (1)$$

where w is the width of the channel, h is Planck's constant, g is the Landé g -factor, μ_B is the Bohr magneton and v_F is the Fermi velocity determined by the density of two-dimensional electrons, n_s : $v_F = \hbar\sqrt{2\pi n_s}/m^*$ with m^* the effective electron mass. Similar to conventional continuous wave electron spin resonance, this resonance leads to a rapid randomization of spin direction and a suppression of spin relaxation length in the channel¹².

We detect this ballistic spin resonance (BSR) by injecting electrons into a narrow channel of high mobility ($\mu = 4.44 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at $n_s = 1.1 \times 10^{11} \text{ cm}^{-2}$ and temperature $T = 1.5 \text{ K}$) GaAs/AlGaAs two-dimensional electron gas (2DEG) through a spin-selective quantum point contact (QPC) (Fig. 1a, b)^{10,13,14}. The charge current is drained at the left end of the channel, while diffusion of the accumulated spin polarization generates a pure spin current to the right of the injector. The spin current can be quantified by the non-local voltage, V_{nl} , that develops across a spin-selective detector QPC located 7–20 μm to the right of the injector, in the region where no charge current flows^{15,16}. The mean free path in the channels, ℓ , significantly exceeds the channel

width ($\ell = 4\text{--}20 \mu\text{m}$ for $w = 1 \mu\text{m}$, and $\ell = 20 \mu\text{m}$ for $w = 3 \mu\text{m}$), so electrons cross the channel several times before their momentum is randomized. In this measurement, the relaxation length of the spin current ranges from tens of micrometres away from resonance to just a few micrometres on resonance.

An external magnetic field applied in the plane of the electron gas, $B_{x(y)}^{\text{ext}}$, serves two purposes. First, it breaks the spin degeneracy of one-dimensional conductance sub-bands in the injector and detector QPCs, setting the direction and magnitude of QPC polarization. Second, it defines the quantization axis and precession frequency for spins as they travel through the channel. Spin dynamics in a GaAs/AlGaAs 2DEG are governed by the total magnetic field, including both the external field and the effective field associated with spin-orbit interaction, $\mathbf{B}^{\text{tot}}(\mathbf{k}) = \mathbf{B}^{\text{ext}} + \mathbf{B}^{\text{so}}(\mathbf{k})$. The spin-orbit field is dominated by first-order contributions arising from bulk (Dresselhaus, β) and structural (Rashba, α) inversion asymmetry^{17,18}:

$$\frac{1}{2} g\mu_B \mathbf{B}^{\text{so}}(\mathbf{k}) = (\alpha - \beta)k_y \hat{x} - (\alpha + \beta)k_x \hat{y} \quad (2)$$

where \hat{x} and \hat{y} define the 2DEG plane and correspond to $[110]$ and $[\bar{1}10]$ crystal axes, respectively.

Qualitatively, equation (2) shows that motion in the \hat{y} direction induces an effective spin-orbit field in the \hat{x} -direction, B_x^{so} ; motion in the \hat{x} direction induces B_y^{so} . As shown in Fig. 1c, electron trajectories that connect the injector with the detector consist of rapid bouncing along \hat{y} leading to a periodic B_x^{so} , while diffusive motion along \hat{x} gives slower random changes in B_y^{so} . Spin resonance requires a periodic field transverse to the external field. The lack of periodicity in B_y^{so} implies no resonance when \mathbf{B}^{ext} is applied along \hat{x} , and indeed none was observed. Instead, the non-local spin signal increases monotonically with B_x^{ext} , reflecting the increase in QPC polarization (Fig. 1c)¹⁰.

When \mathbf{B}^{ext} is applied along \hat{y} , however, the periodic field B_x^{so} leads to BSR and a collapse of the non-local signal at the field $B^{\text{tot}} = B_0$ of equation (1) (Fig. 1c). The centre of the resonance dip near $B_y^{\text{ext}} = 7 \text{ T}$ implies an electron density of $n_s \approx 1 \times 10^{11} \text{ cm}^{-2}$ that is close to the bulk value, and consistent with Hall measurements in the channel. The signal disappears completely in a 1- μm -wide channel at $B_y^{\text{ext}} = 6\text{--}8 \text{ T}$ (Fig. 1c), indicating that the spin current has completely relaxed before reaching the detector QPC 20 μm away.

Measurements of the spin relaxation length confirm that the non-monotonic B_y^{ext} dependence of the spin signal is due to spin relaxation inside the channel (Fig. 1d). The spin relaxation length is deduced from the change in the spin signal as the channel length is adjusted *in situ*¹⁰. The channel can be shortened by undepleting the Λ -gate (Fig. 1b) to the right of the detector. If the relaxation length is long, the spin signal drops when the channel is shortened. In contrast, the spin signal does not change if the spin current relaxes before reaching the Λ -gate (see Supplementary Information for details). The extracted spin relaxation length was $\lambda_s = 50 \pm 10 \mu\text{m}$ for \hat{x} -oriented spins, nearly independent of field over $B_x^{\text{ext}} = 2\text{--}10 \text{ T}$. For fields applied along \hat{y} , on the other hand, λ_s collapsed above

¹Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ²Institut für Angewandte und Experimentelle Physik, Universität Regensburg, Regensburg 93040, Germany.

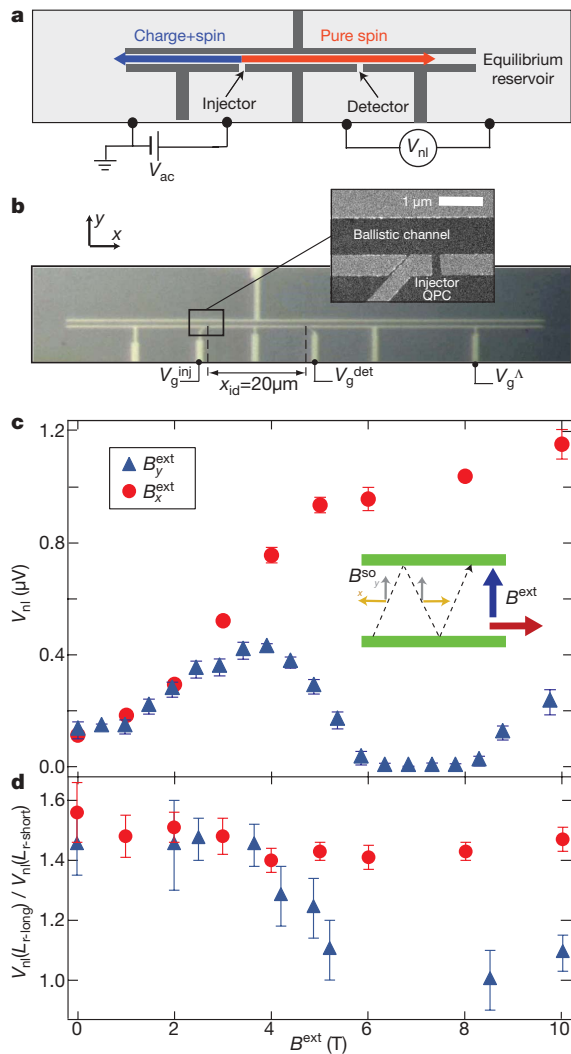


Figure 1 | Ballistic spin resonance. **a**, Simplified diagram of a pure spin current measurement. Gates (dark grey) deplete the 2DEG to define the injector and detector QPCs and the ballistic channel. **b**, Optical image of a measured device with CrAu gates in light grey. Gate voltages V_{inj}^g and V_{det}^g tune QPCs to the spin-polarized plateaus. Undepleting the Λ -gate (V_{Λ}^g) changes the distance from the injector to the right reservoir from $L_{r-long} = 70 \mu m$ to $L_{r-short} = 40 \mu m$. Inset, scanning electron micrograph of the injector area. **c**, Magnetic field dependence of V_{nl} measured with injector and detector set to conductances $G_{inj} = G_{det} \approx 1e^2/h$, for B_y^{ext} applied along \hat{x} and \hat{y} ($w = 1 \mu m$, $T = 300$ mK, $V_{ac} = 50 \mu V$, $x_{id} = 20 \mu m$, $B_z^{ext} = 10$ mT). Finite V_{nl} near $B_y^{ext} = 0$ is due to thermoelectric Peltier effect. Inset, B_x^{so} oscillates, while B_y^{so} is constant along an ideal specular trajectory¹⁰. **d**, Magnetic field dependence of the ratio $V_{nl}(L_{r-long})/V_{nl}(L_{r-short})$ for B_y^{ext} along \hat{x} and \hat{y} . Error bars indicate the standard deviation in the measurements, originating from universal fluctuations of non-local conductance.

$B_y^{ext} = 4$ T, then partially recovered for $B_y^{ext} > 8$ T. The absence of a measurable spin signal for $B_y^{ext} = 6$ –8 T and for the injector–detector spacing $x_{id} = 20 \mu m$ implies an upper bound $\lambda_s < 5 \mu m$ within this range of fields.

Resonant spin dynamics in ballistic channels are influenced by the details of electron trajectories; varying the parameters of typical trajectories changes the resonance conditions. For example, a lower electron density leads to a lower Fermi velocity, hence a lower resonant frequency f_0 (equation (1)), so the BSR dip appears at a lower field B_0 . As the density was lowered in a 1- μm -wide channel (Fig. 2a), B_0 shifted from 7 T down to 4 T. The magnetic field dependence of the spin signal is closely matched by Monte Carlo simulations of spin dynamics in the channel (Fig. 2b; see Supplementary Information for simulation details)¹².

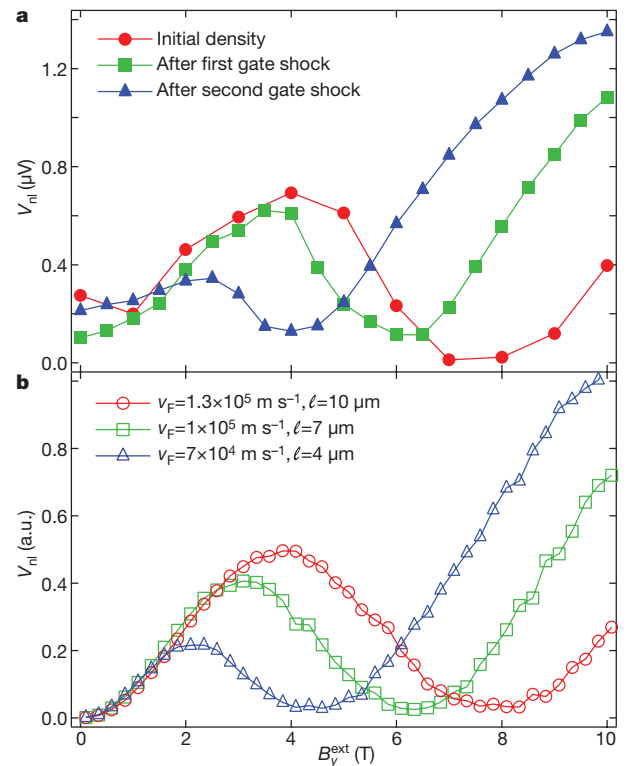


Figure 2 | Tuning the ballistic spin resonance with electron density. **a**, Magnetic field B_y^{ext} dependence of the spin signal for three electron densities in a 1- μm -wide channel. $T = 500$ mK, $V_{ac} = 50 \mu V$, $x_{id} = 6.7 \mu m$, $B_z^{ext} = 40$ mT. Data for the initial density are multiplied by 3 for better comparison to other densities (much lower channel resistance at this density led to a lower non-local voltage). Density was changed by shocking the 2DEG with high negative voltages on the gates (see Supplementary Information for details). **b**, Monte Carlo simulations of the spin signal for three Fermi velocities, v_F (in m s⁻¹): 1.3×10^5 , 1×10^5 and 7×10^4 .

Higher frequency harmonics of the effective spin–orbit field can also be expected. Bouncing off the channel walls leads to a square wave dependence for the velocity component $v_y(t) \propto B_x^{so}(t)$, with Fourier components at $(2N + 1)f_0$ (Fig. 3b–d). In principle, this leads to a ladder of BSRs extending to arbitrarily high external magnetic fields, but in practice, the higher harmonics disappear quickly because of scattering. A dip in the spin signal was observed in a wider channel ($w = 3 \mu m$) at a field $B_0 = 8$ T that corresponds to the third harmonic $3f_0$ (Fig. 3a). The first harmonic BSR dip in a 3- μm -wide channel is expected at $B_0 = 2.6$ T from simulations. QPC polarization is too small at this field to clearly resolve the first resonance. Instead, the primary visible effect of BSR is to counteract the increase in QPC polarization with magnetic field, giving a flatter spin signal in the range $B_y^{ext} = 0$ –5 T compared to the 1- μm -wide channel (Fig. 1c versus Fig. 3a).

The third harmonic resonance can be clearly seen in the Monte Carlo simulation if one assumes a field independent QPC polarization and long mean free path $\ell = 30 \mu m$ (Fig. 3a Sim A). It is remarkable that simulated trajectories even with such a long mean free path look so disordered to the eye (Fig. 3b), but the resilience of the resonance to small amounts of disorder can be understood qualitatively from the square-wave character of the velocity (Fig. 3c). More realistic simulation parameters, with shorter mean free path and finite perpendicular field, provide a better match to the measured resonances (Fig. 3a Sim B).

A small component of magnetic field applied perpendicular to the 2DEG, $B_z^{ext} \approx 10$ –100 mT ($\ll B_{xy}^{ext}$), bends electron trajectories in partial cyclotron orbits, generating a qualitatively different spectrum of spin–orbit fields (Fig. 4). Cyclotron trajectories contain oscillating components along both \hat{x} and \hat{y} , giving rise to periodic spin–orbit

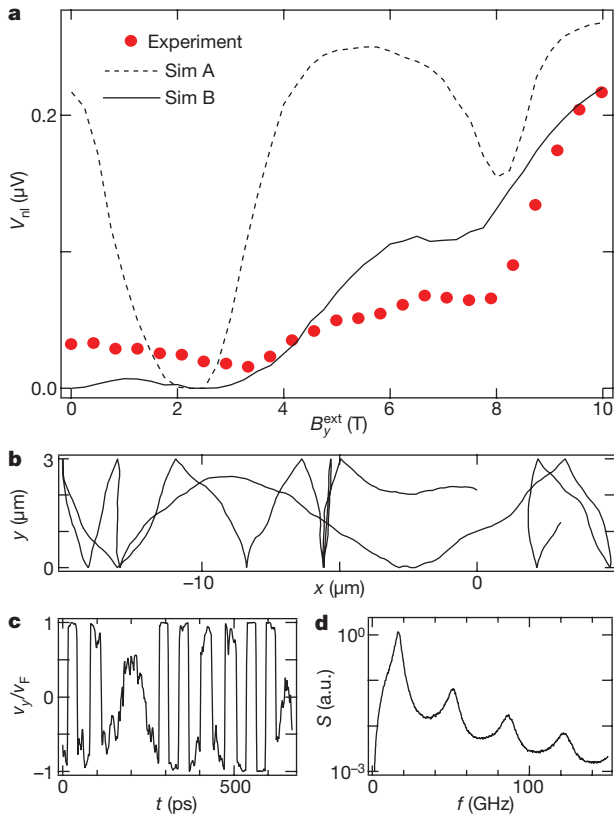


Figure 3 | Third harmonic resonance. **a**, Measured (circles) and simulated (lines) spin signal in a 3- μm -wide channel ($T = 500\text{ mK}$, $V_{ac} = 50\text{ }\mu\text{V}$, $x_{id} = 6.7\text{ }\mu\text{m}$, $B_z^{ext} = 20\text{ mT}$). The non-local voltage plotted is the difference between signals for spin-polarized and spin-unpolarized detector gate settings to reduce spurious signals not due to spin. Simulation (Sim) A, simulation parameters tuned to show two dips clearly. $v_F = 1.1 \times 10^5\text{ m s}^{-1}$, $\ell = 30\text{ }\mu\text{m}$, $B_z^{ext} = 0\text{ mT}$ and a constant QPC polarization. Sim B, realistic parameters approximate data: $v_F = 1.1 \times 10^5\text{ m s}^{-1}$, $\ell = 10\text{ }\mu\text{m}$, $B_z^{ext} = 20\text{ mT}$. **b**, Simulated segment of an electron trajectory in a 3- μm -wide channel (parameters as in Sim A above). **c**, The component of electron velocity v_y for the trajectory in **b** is reminiscent of a square wave, changing sign each $\approx 30\text{ ps}$ (the crossing time of the channel). **d**, Power spectrum of v_y shows a comb of peaks that correspond to BSR at odd multiples of the primary frequency $f_0 = 16.7\text{ GHz}$.

fields $B_y^{so}(t)$ that were not present at $B_z^{ext} = 0$ (Fig. 4b). A periodic $B_y^{so}(t)$ causes BSR when the in-plane external field is applied along \hat{x} (Fig. 4a). This is in sharp contrast to the case of small perpendicular field B_z^{ext} , where BSR was robustly absent in B_x^{ext} in many samples and cooldowns (Fig. 1c)¹⁰. When the perpendicular field makes the cyclotron radius less than the channel width, $r_c \equiv m^*v_F/eB_z^{ext} < w$, the characteristic bouncing frequency increases as electrons follow skipping orbits that do not extend across the entire channel (Fig. 4c). This leads to an increase in the resonance field B_0 observed in the measurement and matched by simulation.

The qualitative effects of B_z^{ext} can be seen in Fourier analyses of electron momenta along \hat{x} and \hat{y} for realistic trajectories (Fig. 4d, e). The spectrum of $v_x(t)$ shows no finite-frequency peak at $B_z^{ext} = 0$, but a spectral component begins to emerge at 45 GHz above $B_z^{ext} \approx 10\text{ mT}$ leading to the B_x^{ext} spin resonance in Fig. 4a. Above $B_z^{ext} \approx 40\text{ mT}$ the cyclotron radius falls below the channel width, so most trajectories do not cross the channel and the resonant frequency approaches the cyclotron frequency as the field is increased. Higher harmonics are quickly suppressed owing to a transition from switching (square wave) to smooth changes in $B_x^{so}(t)$.

Equation (2) shows that a partial cancellation of bulk and structural asymmetry terms in the spin-orbit interaction may lead to different values of B_x^{so} and B_y^{so} . Figures 2 and 3 show BSR for B_y^{ext} , which depends

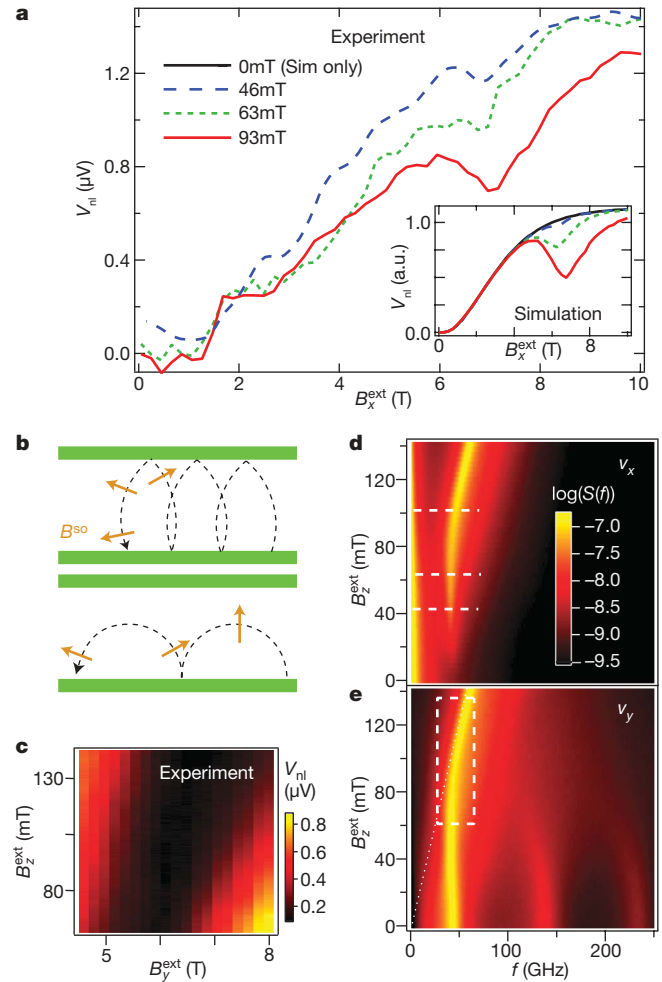


Figure 4 | The effect of out-of-plane magnetic field. **a**, Resonance emerges for $B_z^{ext} \parallel \hat{x}$ when finite B_z^{ext} bends trajectories, giving rise to periodicity in $B_y^{so}(t)$. Inset, simulation for realistic parameters ($v_F = 1 \times 10^5\text{ m s}^{-1}$, $\ell = 10\text{ }\mu\text{m}$). **b**, Cartoons of trajectories for cyclotron radius greater than (upper panel) and less than (lower panel) the channel width. Effective spin-orbit fields (orange arrows) oscillate along both \hat{x} and \hat{y} . **c**, Resonance shifts to higher magnetic field B_0 when cyclotron orbits do not cross the channel, shown here for $B_z^{ext} \parallel \hat{y}$. **d**, **e**, Simulated power spectra for v_x and v_y with parameters from the inset of **a**. Dashed lines in **d** and **e** show range of experimental data in **a** and **c**, respectively. Dotted line in **e** is twice the cyclotron frequency. Colour scale denotes logarithm of spectral power. Throughout this figure: $w = 1\text{ }\mu\text{m}$, $T = 500\text{ mK}$, $V_{ac} = 50\text{ }\mu\text{V}$, $x_{id} = 6.7\text{ }\mu\text{m}$.

primarily on $B_x^{so} \propto (\alpha - \beta)k_F$. The resonance in B_x^{ext} , in Fig. 4a, depends primarily on $B_y^{so} \propto (\alpha + \beta)k_F$. A comparison of these data sets with simulation enables an estimation of the spin-orbit anisotropy: the best match was found for $|\alpha - \beta| = (2 \pm 0.5) \times 10^{-13}\text{ eV m}$ and $|\alpha + \beta| = (0.6 \pm 0.2) \times 10^{-13}\text{ eV m}$. These values were used in the simulations throughout this Letter.

Results presented here demonstrate that the effect of spin-orbit interaction on spin coherence can be dramatically amplified with a geometric resonance. In quasi-diffusive structures like the channels shown here, spin polarization is randomized at the resonance. Coherent spin rotations could be achieved by the same mechanism if the effects of scattering were small and the number of ballistic cycles were well-defined, as is the case in the transverse electron focusing geometry¹³.

METHODS SUMMARY

Devices were defined in a triangular well 2D electron gas 110 nm under the surface of a [001] GaAs/AlGaAs heterostructure, using CrAu electrostatic gates deposited on the surface. Voltage biased lock-in measurements in a dilution refrigerator were performed in magnetic fields, B^{ext} , up to 10 T in the plane of

the 2DEG. The out-of-plane component, B_z^{ext} , could be controlled independently. Data presented here are from 3 channels of length 100 μm . Electron density in Fig. 2 was changed by applying voltage shocks to the gates. Spin dynamics along random trajectories were calculated using the spin–orbit field given by equation (2), and assuming realistic 2DEG parameters. Spin relaxation length extracted from the simulations was substituted into the spin diffusion equation and multiplied by the magnetic-field-dependent QPC polarization to obtain a simulated spin signal at the detector.

Received 6 November 2008; accepted 6 February 2009.

1. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A new method of measuring nuclear magnetic moment. *Phys. Rev.* **53**, 318 (1938).
2. Zavoisky, Y. K. Spinmagnetic resonance in paramagnetics. *J. Phys. USSR* **9**, 245–246 (1945).
3. Dyakonov, M. I. & Perel, V. I. Spin relaxation of conduction electrons in noncentrosymmetric semiconductors. *Sov. Phys. Solid State* **13**, 3023–3026 (1972).
4. Bell, R. L. Electric dipole spin transitions in InSb. *Phys. Rev. Lett.* **9**, 52–54 (1962).
5. Rashba, E. I. & Efros, A. L. Orbital mechanisms of electron-spin manipulation by an electric field. *Phys. Rev. Lett.* **91**, 126405 (2003).
6. Duckheim, M. & Loss, D. Electric-dipole-induced spin resonance in disordered semiconductors. *Nature Phys.* **2**, 195–199 (2006).
7. Kato, Y., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Coherent spin manipulation without magnetic fields in strained semiconductors. *Nature* **427**, 50–53 (2004).
8. Meier, L. *et al.* Measurement of Rashba and Dresselhaus spin-orbit magnetic fields. *Nature Phys.* **3**, 650–654 (2007).
9. Nowack, K. C., Koppens, F. H. L., Nazarov, Y. V. & Vandersypen, L. M. K. Coherent control of a single electron spin with electric fields. *Science* **318**, 1430–1433 (2007).
10. Frolov, S. M. *et al.* Electrical generation of pure spin currents in a two-dimensional electron gas. *Phys. Rev. Lett.* **102**, 116802 (2009).
11. Wolf, S. A. *et al.* Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
12. Koop, E. J., van Wees, B. J. & van der Wal, C. H. Confinement-enhanced spin relaxation for electron ensembles in large quantum dots. Preprint at (<http://arxiv.org/abs/0804.2968>) (2008).
13. Potok, R. M., Folk, J. A., Marcus, C. M. & Umansky, V. Detecting spin-polarized currents in ballistic nanostructures. *Phys. Rev. Lett.* **89**, 266602 (2002).
14. Koop, E. J. *et al.* Spin accumulation and spin relaxation in a large open quantum dot. *Phys. Rev. Lett.* **101**, 056602 (2008).
15. Johnson, M. & Silsbee, R. H. Interfacial charge-spin coupling — Injection and detection of spin magnetization in metals. *Phys. Rev. Lett.* **55**, 1790–1793 (1985).
16. Jedema, F. J., Filip, A. T. & van Wees, B. J. Electrical spin injection and accumulation at room temperature in an all-metal mesoscopic spin valve. *Nature* **410**, 345–348 (2001).
17. Rashba, E. I. Properties of semiconductors with an extremum loop. *Sov. Phys. Solid State* **2**, 1224–1238 (1960).
18. Dresselhaus, G. Spin-orbit coupling effects in zinc blende structures. *Phys. Rev.* **100**, 580–586 (1955).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Berciu, M. Duckheim, J. C. Egues, D. Loss and G. Usaj for conversations. Work at UBC was supported by NSERC, CFI and CIFAR. W.W. acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the programme 'Halbleiter-Spintronik' (SPP 1285).

Author Contributions S.M.F., J.A.F., W.Y. and Y.R. performed experiments, S.L. performed Monte Carlo simulations, W.W. provided heterostructures, and S.M.F. and J.A.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.F. (jfolk@physics.ubc.ca).

LETTERS

Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons

Dmitry V. Kosynkin¹, Amanda L. Higginbotham¹, Alexander Sinitskii¹, Jay R. Lomeda¹, Ayrat Dimiev¹, B. Katherine Price¹ & James M. Tour^{1,2,3}

Graphene, or single-layered graphite, with its high crystallinity and interesting semimetal electronic properties, has emerged as an exciting two-dimensional material showing great promise for the fabrication of nanoscale devices^{1–3}. Thin, elongated strips of graphene that possess straight edges, termed graphene ribbons, gradually transform from semiconductors to semimetals as their width increases^{4–7}, and represent a particularly versatile variety of graphene. Several lithographic^{7,8}, chemical^{9–11} and synthetic¹² procedures are known to produce microscopic samples of graphene nanoribbons, and one chemical vapour deposition process¹³ has successfully produced macroscopic quantities of nanoribbons at 950 °C. Here we describe a simple solution-based oxidative process for producing a nearly 100% yield of nanoribbon structures by lengthwise cutting and unravelling of multiwalled carbon nanotube (MWCNT) side walls. Although oxidative shortening of MWCNTs has previously been achieved¹⁴, lengthwise cutting is hitherto unreported. Ribbon structures with high water solubility are obtained. Subsequent chemical reduction of the nanoribbons from MWCNTs results in restoration of electrical conductivity. These early results affording nanoribbons could eventually lead to applications in fields of electronics and composite materials where bulk quantities of nanoribbons are required^{15–17}.

We obtained oxidized nanoribbons by suspending MWCNTs in concentrated sulphuric acid followed by treatment with 500 wt% KMnO₄ for 1 h at room temperature (22 °C) and 1 h at 55–70 °C (Methods). After isolation, the resulting nanoribbons were highly soluble in water (12 mg ml^{–1}), ethanol and other polar organic solvents. The opening of the nanotubes appears to occur along a line, similar to the ‘unzipping’ of graphite oxide^{18,19}, affording straight-edged ribbons. This could occur in a linear longitudinal cut (Fig. 1a) or in a spiralling manner, depending upon the initial site of attack and the chiral angle of the nanotube. Although depicted in Fig. 1a as occurring on the mid-section of the nanotube rather than at one end, the location of the initial attack is not known.

The mechanism of opening is based on previous work on the oxidation of alkenes by permanganate in acid. The proposed first step in the process is manganate ester formation (2, Fig. 1b) as the rate-determining step, and further oxidation is possible to afford the dione (3, Fig. 1b) in the dehydrating medium²⁰. Juxtaposition of the buttressing ketones distorts the β,γ -alkenes (red in 3), making them more prone to the next attack by permanganate. As the process continues, the buttressing-induced strain on the β,γ -alkenes lessens because there is more space for carbonyl projection; however, the bond-angle strain induced by the enlarging hole (or tear if originating from the end of the nanotube) would make the β,γ -alkenes (4, Fig. 1b) increasingly reactive. Hence, once an opening has been initiated, its further opening is enhanced relative to an unopened tube or to an

uninitiated site on the same tube. The ketones can be further converted, through their O-protonated forms, to the carboxylic acids²¹ that will line the edges of the nanoribbons. Finally, relief of the bond-angle strain when the nanotube opens to the graphene ribbon (5, Fig. 1b) slows further dione formation and cutting²⁰. Thus, the preference for sequential bond cleavage over random opening and subsequent cutting, as occurs with nitric acid oxidation, can be explained by concerted attachment to neighbouring carbon atoms by permanganate, contrasting with the random attack on non-neighbouring carbon atoms by the nitronium species from nitric acid. The surface of the now-less-strained nanoribbon remains prone to 1,2-diol formation, which leads to the overall highly oxidized ribbon, but this is less likely to result in further oxidative cutting to the dione owing to relief of the tubular strain on the double bonds.

We achieved the same unzipping process in single-walled carbon nanotubes (SWCNTs), to produce narrow nanoribbons, but their subsequent disentanglement is more difficult. See Supplementary Figs 5 and 6 for images and analysis of those SWCNT-derived narrow nanoribbons and their reduction products.

We used transmission electron microscopy (TEM), atomic force microscopy (AFM) and scanning electron microscopy (SEM) to image the ribbon structures. TEM analysis shows nanoribbons (Fig. 1c) produced from MWCNTs with a starting diameter of 40–80 nm and approximately 15–20 inner nanotube layers (additional TEM images of untreated MWCNTs can be found in the Supplementary Fig. 1a, b). After reaction, the width of the carbon nanostructures increased to >100 nm and they had linear edges with little pristine MWCNT side-wall structure remaining (see Supplementary Figs 1 and 2 for more images). The MWCNTs used were produced from a chemical vapour deposition process²²; we attempted the same H₂SO₄–KMnO₄ treatment on a single sample of laser-oven-produced MWCNTs, but fewer nanoribbon-like structures were detected. AFM imaging (Fig. 1d) shows the presence of single atomic layers after tip sonication of the solution for 30 min to yield well-dispersed and sonication-shortened ribbons suitable for imaging. SEM imaging (Fig. 1e) of nanoribbons on a silicon surface shows that the ribbons remain long (~4 μ m in this image) when not cut by tip sonication; they can be dispersed as single or thin layers and they display uniform widths and predominantly straight edges over their entire length (see Supplementary Fig. 1c, d for other images).

The degree of consecutive tube opening in the MWCNTs can also be controlled by adjusting the amount of oxidizing agent introduced into the system; using TEM, we found that in 80–100% of the MWCNTs present, the side walls completely unravelled to form nanoribbons when 500 wt% KMnO₄ was used. The successive opening reaction was demonstrated in five iterations, each containing a stepwise increase in the amount of KMnO₄: 100 wt% KMnO₄ in the

¹Department of Chemistry, ²Department of Mechanical Engineering and Materials Science, ³Smalley Institute for Nanoscale Science and Technology, Rice University, MS-222, 6100 Main Street, Houston, Texas 77005, USA.

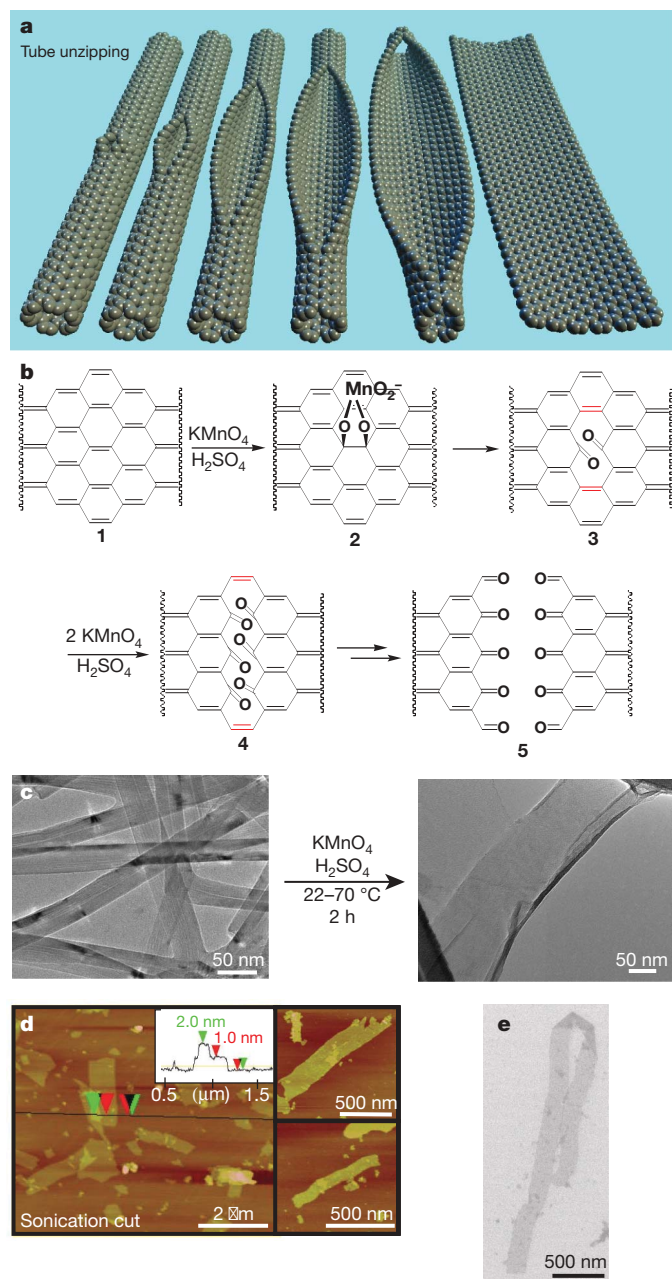


Figure 1 | Nanoribbon formation and imaging. **a**, Representation of the gradual unzipping of one wall of a carbon nanotube to form a nanoribbon. Oxygenated sites are not shown. **b**, The proposed chemical mechanism of nanotube unzipping. The manganate ester in **2** could also be protonated. **c**, TEM images depicting the transformation of MWCNTs (left) into oxidized nanoribbons (right). The right-hand side of the ribbon is partly folded onto itself. The dark structures are part of the carbon imaging grid. **d**, AFM images of partly stacked multiple short fragments of nanoribbons that were horizontally cut by tip-ultrasonic treatment of the original oxidation product to facilitate spin-casting onto the mica surface. The height data (inset) indicates that the ribbons are generally single layered. The two small images on the right show some other characteristic nanoribbons. **e**, SEM image of a folded, 4-μm-long single-layer nanoribbon on a silicon surface.

first iteration (sample I), 200 wt% in the second iteration (sample II), and so on until the final iteration, when we used 500 wt% (sample V). This resulted in consecutive unencapsulation of the different layers by unzipping of the successive MWCNTs (see Methods for details). It is evident from TEM images (Fig. 2a–e) that the walls of the MWCNTs open to a higher degree as the level of oxidation increases, with less MWCNT inner tube remaining in successive iterations. This is highlighted in a statistical plot (Fig. 2f) showing the decrease of the

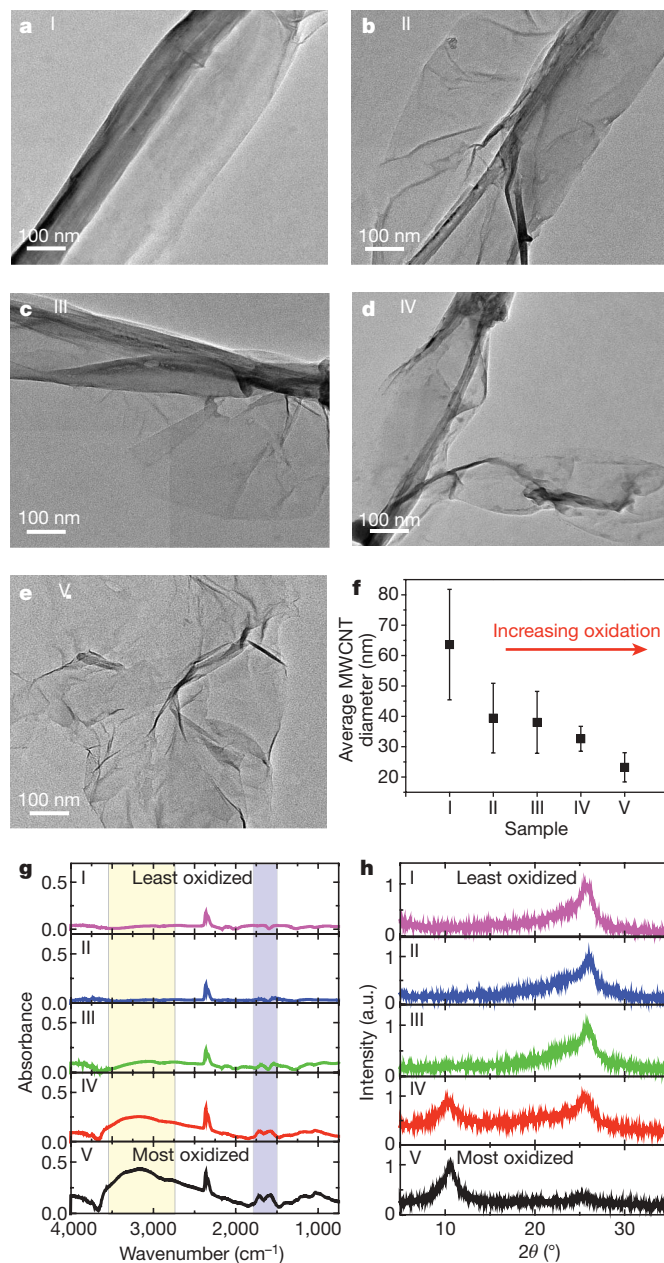


Figure 2 | Stepwise opening of MWCNTs to form nanoribbons. **a–e**, TEM images of the stepwise opening of MWCNTs representing the incremental exposure of the system to KMnO_4 ; the least oxidized sample (sample I) is in **a** and the most oxidized sample (sample V) is in **e**. **f**, Scatter plot showing how the average MWCNT diameter (determined from studying 15–20 TEM images per sample, each with ~5 MWCNTs per image) changes with increasing exposure to KMnO_4 . Error bars indicate the standard deviation of the average MWCNT diameter across the sample. **g**, ATR-IR spectroscopy of stepwise opening/oxidation of MWCNTs. **h**, X-ray diffraction analysis of the stepwise opening of the nanotube. θ , diffraction angle; a.u., arbitrary units.

average diameter of remaining MWCNTs from ~65 nm to ~20 nm as the amount of KMnO_4 exposure is increased. The smaller-diameter tubes that remained after treatment with 500 wt% KMnO_4 were exposed to the reaction conditions for less time than the larger-diameter tubes and, thus, may not have had the chance to fully react; no difference in the rate of unzipping between smaller- and larger-diameter nanotubes can be inferred from this data.

The degree of oxidation of the product formed (partly and/or completely unravelled MWCNTs) from each of the five iterative KMnO_4 treatment steps was monitored using attenuated-total-reflection infrared (ATR-IR) spectroscopy and thermogravimetric

analysis (TGA). ATR-IR spectroscopy (Fig. 2g) reveals the appearance of a C=O stretch (purple region, Fig. 2g) increasing from $1,690\text{ cm}^{-1}$ in sample III (green line) to $1,710\text{ cm}^{-1}$ in sample V (black line), consistent with declining conjugation. The COO-H/O-H stretch ($\sim 3,600\text{--}2,800\text{ cm}^{-1}$; yellow region, Fig. 2g) appears with sample III and continues to increase through the series, indicating an increase in the number of carboxyl and hydroxyl functionalities as well as the possible presence of trapped water. TGA shows an increase in the total weight loss (20% and 49% in samples I and V, respectively) with increasing exposure to KMnO_4 , implying an increase in the number of volatile side-wall functionalities present, which corroborates there being a higher degree of oxidation (Supplementary Fig. 3a). Furthermore, Raman spectroscopy (Supplementary Fig. 3b) shows an increasing level of disorder (appearance of a D band at $1,321\text{--}1,328\text{ cm}^{-1}$) with increasing oxidation, consistent with ATR-IR spectroscopy and TGA observations.

We also performed X-ray diffraction analysis (Fig. 2h), to investigate further the structure of the partly and completely unzipped MWCNT-nanoribbon structures. The graphite (002) spacing increases with the level of oxidation. Samples I–III all have 2θ values of $\sim 25.8^\circ$, corresponding to a d spacing of 3.4 \AA . Sample IV shows two peaks, one at 10.8° and one at 25.4° , with d spacings of 8.2 \AA and 3.5 \AA , respectively. Sample V shows a predominant peak at 10.6° , corresponding to a d spacing of 8.3 \AA , with minimal signal contributed by MWCNTs ($2\theta = 25.8^\circ$); this spectrum is very similar to that of graphite oxide (Supplementary Fig. 3c).

Both the nanoribbons and graphite oxide possess oxygen-containing functionalities such as carbonyls, carboxyls and hydroxyls²³ that have been shown to exist at the edges and the surface²⁴. The surface oxidation disrupts the π -conjugated network, rendering the nanoribbons and graphite oxide poorly conductive. Hydrazine (N_2H_4) reduction of graphite oxide²⁵ is known to provide a means of restoring conjugation and, thus, some of the conductivity, to form chemically converted graphene (CCG)^{25–28}. The structure of CCG is thought to be a patchwork of intact graphene islands interspersed with regions of tetrahedral sp^3 -hybridized carbon atoms due to incomplete reduction and incomplete re-aromatization; therefore, the electrical conductivity is not as high as that found in the original graphite²⁶. The carboxyl groups, which are found predominately at the edges²⁵, are not reduced by N_2H_4 and remain in the product, further disrupting the π network^{7,8}. Furthermore, as the number of oxygen-containing functionalities decreases during the reduction process, the tendency to aggregate as a result of π stacking increases.

The reduction of oxidized nanoribbons was carried out with aqueous N_2H_4 in the presence of ammonia. To prevent re-aggregation during the reduction procedure, we first dispersed the nanoribbons in an

aqueous surfactant solution, sodium dodecyl sulphate (SDS), to produce stable dispersions of reduced nanoribbons that retained their straight-edged structure (Fig. 3a). The reaction progress was monitored by ultraviolet absorption (Fig. 3b); the bathochromic shift of λ_{max} and the hyperchromicity over the entire range ($>230\text{ nm}$) indicates that electronic conjugation of the ribbons was restored²⁵.

To provide further evidence that the reduction procedure decreased the number of oxygen-containing functionalities from the nanoribbon surface, we performed ATR-IR spectroscopy, X-ray photoemission spectroscopy (XPS; Fig. 3c, d) and TGA. The reduced nanoribbons show almost complete elimination of the COO-H/O-H stretching region ($\sim 3,600\text{--}2,800\text{ cm}^{-1}$; yellow region, Fig. 3c) and a significant decrease in the C=O stretching region ($\sim 1,710\text{ cm}^{-1}$; purple region, Fig. 3c) in the ATR-IR spectrum (blue line, Fig. 3c) in comparison with the intense COO-H/O-H and C=O stretches observed for the oxidized nanoribbons (red line, Fig. 3c). Edge carboxylic acids will remain.

In the XPS carbon 1s spectra of the oxidized and reduced nanoribbons (Fig. 3d), the signals at 286 eV and 287 eV correspond to C–O and C=O, respectively. The shoulder at 289 eV is assigned to carboxyl groups. Upon reduction (blue line), the 286- and 287-eV peaks diminish to a shoulder of the C–C peak (284.4 eV), indicating significant deoxygenation of the nanoribbons by N_2H_4 . As reported for CCG, the most dominant peak after reduction is the C–C peak at 284.8 eV (ref. 29). In addition, the XPS-determined atomic concentration of oxygen (complete table found in Supplementary Fig. 4b) decreases from 42% to 16% upon reduction, but is still higher than the oxygen content of MWCNTs (2.1%), owing, in large part, to the edge carboxylic acid moieties.

The TGA weight loss of the reduced nanoribbons was $\sim 33\%$ less than that of the oxidized starting material, which also indicates that fewer oxygen-containing functionalities are present on the nanoribbon surface (see Supplementary Fig. 4d for TGA curves). The TEM image of a reduced nanoribbon shows its straight edge and buckled appearance (Fig. 3a). Nitrogen adsorption measurements of as-prepared and reduced nanoribbons give surface areas, determined using Brunauer–Emmett–Teller theory, of 445 and $436\text{ m}^2\text{ g}^{-1}$, respectively, after pre-outgassing at 400°C for 12 h (ref. 13). The density of the oxidized ribbons was found to be 2.0 g cm^{-3} using solution density matching (bromotrichloromethane). When considering the overall conversion of MWCNTs to reduced nanoribbons, the material weight yield is 99% (Methods), underscoring the efficiency of the overall process.

Recent interest in graphene nanoribbons has focused on the study of the reactive edges having zigzag or armchair morphologies that dominate their electronic and magnetic behaviour⁵. Although zigzag-edged structures are presumed by the mechanism described

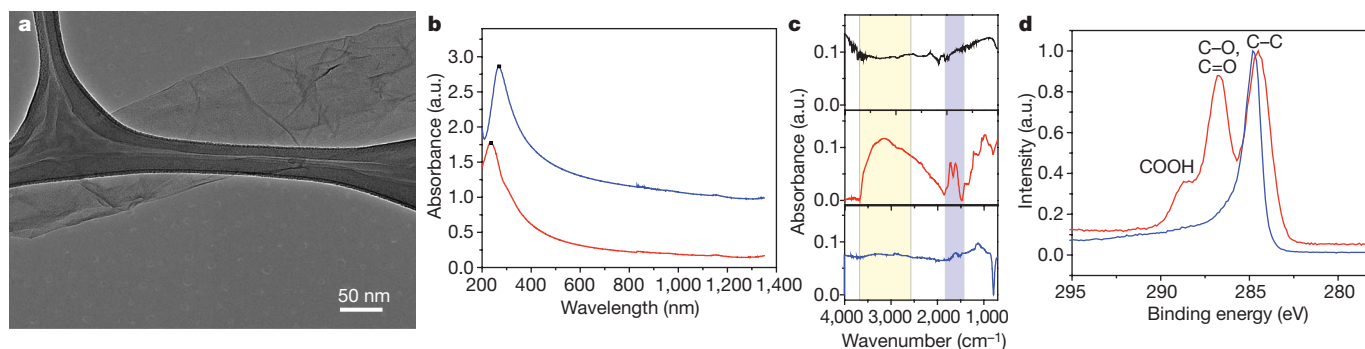


Figure 3 | Characterization of the oxidized and reduced nanoribbons derived from MWCNTs. **a**, TEM image of reduced nanoribbons obtained by treatment of oxidized nanoribbons with N_2H_4 . Detailed examination of the image reveals that 2–3 ribbons are stacked with apparent buckling. The dark structures are part of the carbon imaging grid. **b**, Changes in the ultraviolet spectrum of an aqueous solution of oxidized nanoribbons (red),

$\lambda_{\text{max}} = 234\text{ nm}$) after treatment with N_2H_4 (blue, $\lambda_{\text{max}} = 267\text{ nm}$). **c**, ATR-IR spectroscopy of nanoribbons before (red) and after reduction (blue), compared with MWCNT starting material (black). **d**, Normalized, superimposed XPS carbon 1s spectra of the oxidized nanoribbons (red) and the reduced nanoribbons (blue).

here, we were unable to achieve the edge resolution needed to confirm this. This is due, in part, to edge curling, and could be further attributed to the extensive edge oxidation; this may be removed only upon treatment at $\geq 2,000^\circ\text{C}$, which would result in reconstruction and modified electronic properties.

In spite of the oxidized edges and planes of the nanoribbons derived by this bulk process, the electronic properties of the nanoribbons before and after chemical reduction were studied by building three-terminal devices on Si–SiO₂ substrates. The long length of the nanoribbons make them easily adapted structures for device fabrication (Fig. 4a); electron-beam-patterned platinum electrodes were evaporated on top of the nanoribbon stack. As-prepared nanoribbons are poor conductors owing to the high number of oxygen-containing functionalities present on the surface; however, their conductivity can be dramatically increased either by chemical reduction using N₂H₄ or by annealing in H₂ (Fig. 4b). Thick nanoribbon stacks show little gate effect, which is in accord with previously reported data³⁰. Conversely, bilayers of these reduced graphene nanoribbons have field-effect properties with a minimum conductivity at zero gate voltage, which is as expected for undoped field-effect devices made from exfoliated graphene sheets and is superior to CCGs (Fig. 4c, d)^{2,30}. The conductivities obtained from these wide nanoribbons are analogous to device properties reported^{11,13} for other wide nanoribbons either exfoliated or grown by chemical vapour deposition. We have so far been unable to build acceptable devices from narrow nanoribbons derived from SWCNTs, owing to their extreme entanglement (Supplementary Fig. 5); moreover, edge oxidation in those small structures may retard their electronic utility. Although the preparative route described here can have the advantage of producing accessible nanoribbons on a large scale, these unzipping-derived nanoribbons, with their residual oxidized

defect sites, have electronic characteristics inferior to those of wide, mechanically peeled sheets of graphene^{2,30}.

METHODS SUMMARY

Nanoribbon formation. MWCNTs were used as received from Mitsui & Co. (lot no. 05072001K28). We suspended MWCNTs in concentrated sulphuric acid (H₂SO₄) for a period of 1–12 h and then treated them with 500 wt% potassium permanganate (KMnO₄). The H₂SO₄ conditions aid in exfoliating the nanotube and the subsequent graphene structures. The reaction mixture was stirred at room temperature for 1 h and then heated to 55–70 °C for an additional 1 h. When all of the KMnO₄ had been consumed, we quenched the reaction mixture by pouring over ice containing a small amount of hydrogen peroxide (H₂O₂). The solution was filtered over a polytetrafluoroethylene (PTFE) membrane, and the remaining solid was washed with acidic water followed by ethanol/ether.

Stepwise oxidation of MWCNTs to nanoribbons. We followed the above reaction procedure, except that 100 wt% KMnO₄ was added in portions until 500 wt% was achieved. When the KMnO₄ had been consumed at every step, a portion of the reaction solution was removed and worked up for analysis as described above.

Nanoribbon reduction. We treated a water solution (200 mg l⁻¹) of the above-isolated nanoribbons (with or without 1 wt% SDS surfactant) with 1 vol% concentrated ammonium hydroxide (NH₄OH) and 1 vol% hydrazine monohydrate (N₂H₄·H₂O). Before being heating to 95 °C for 1 h, the solution was covered with a thin layer of silicon oil.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 October 2008; accepted 11 February 2009.

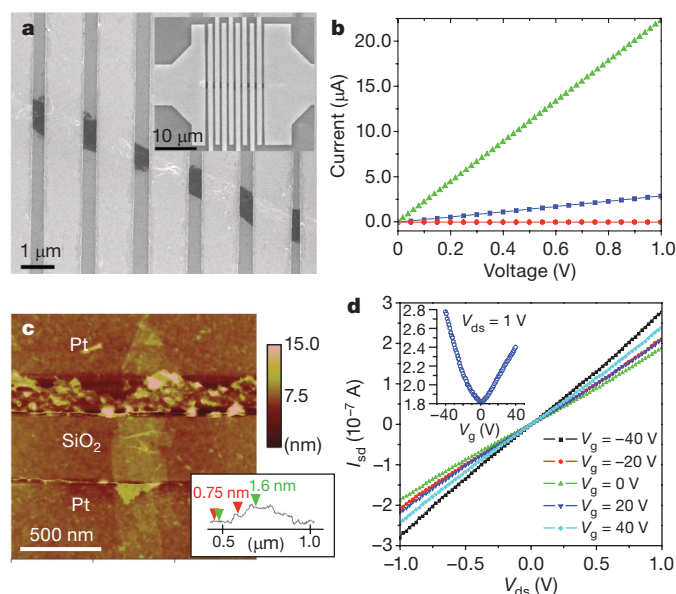


Figure 4 | Device fabrication and electrical properties of graphene nanoribbons on SiO₂-Si. **a**, SEM image of a multi-terminal device based on a multilayer (~ 10 -nm-thick) stack of graphene nanoribbons with platinum electrodes. Inset, larger image of a similar device. **b**, Current–voltage curves for three different types of device: as-prepared (red), N₂H₄-reduced (blue) and H₂-annealed (green) nanoribbons (~ 300 nm wide, 10 nm thick (AFM) with a channel length of ~ 500 nm; characteristic of the > 10 devices measured for each of the three states). **c**, AFM image of another device based on a N₂H₄-reduced and annealed (H₂/Ar at 300 °C for 10 min) bilayer nanoribbon showing that the ribbon consists of two overlapping nanoribbons in the SiO₂ channel region. Typical height profile (inset) across this nanoribbon shows steps of about 0.75 nm, which correspond to individual graphene sheets. These sheets overlap in the middle, resulting in a height of ~ 1.6 nm. **d**, Source–drain current (I_{sd}), source–drain voltage (V_{ds}) and gate voltage (V_g) dependencies for the device shown in **c**; p-doped silicon was used as a back gate.

- Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).
- Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Areshkin, D. A., Gunlycke, D. & White, C. T. Ballistic transport in graphene nanostrips in the presence of disorder: importance of edge effects. *Nano Lett.* **7**, 204–210 (2007).
- Nakada, K., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Edge state in graphene ribbons: nanometer size effect and edge shape dependence. *Phys. Rev. B* **54**, 17954–17961 (1996).
- Son, Y.-W., Cohen, M. L. & Louie, S. G. Energy gaps in graphene nanoribbons. *Phys. Rev. Lett.* **97**, 216803 (2006).
- Han, M. Y., Oezylmaz, B., Zhang, Y. & Kim, P. Energy band-gap engineering of graphene nanoribbons. *Phys. Rev. Lett.* **98**, 206805 (2007).
- Chen, Z., Lin, Y.-M., Rooks, M. J. & Avouris, P. Graphene nano-ribbon electronics. *Physica E* **40**, 228–232 (2007).
- Schniepp, H. C. *et al.* functionalized single graphene sheets derived from splitting graphite oxide. *J. Phys. Chem. B* **110**, 8535–8539 (2006).
- Rollings, E. *et al.* Synthesis and characterization of atomically thin graphite films on a silicon carbide substrate. *J. Phys. Chem. Solids* **67**, 2172–2177 (2006).
- Li, X., Wang, X., Zhang, L., Lee, S. & Dai, H. Chemically derived, ultrasmooth graphene nanoribbon semiconductors. *Science* **319**, 1229–1232 (2008).
- Yang, X. *et al.* Two-dimensional graphene nanoribbons. *J. Am. Chem. Soc.* **130**, 4216–4217 (2008).
- Campos-Delgado, J. *et al.* Bulk production of a new form of sp² carbon: crystalline graphene nanoribbons. *Nano Lett.* **8**, 2773–2778 (2008).
- Saito, T., Matsushige, K. & Tanaka, K. Chemical treatment and modification of multi-walled carbon nanotubes. *Physica B* **323**, 280–283 (2002).
- Son, Y.-W., Cohen, M. L. & Louie, S. G. Half-metallic graphene nanoribbons. *Nature* **444**, 347–349 (2006).
- Liang, G., Neophytou, N., Nikonov, D. E. & Lundstrom, M. S. Performance projections for ballistic graphene nanoribbon field-effect transistors. *IEEE Trans. Electron. Dev.* **54**, 677–682 (2007).
- Wang, X. *et al.* Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors. *Phys. Rev. Lett.* **100**, 206803 (2008).
- Li, J.-L. *et al.* Oxygen-driven unzipping of graphitic materials. *Phys. Rev. Lett.* **96**, 176101 (2006).
- Ajayan, P. M. & Yakobson, B. I. Oxygen breaks into carbon world. *Nature* **441**, 818–819 (2006).
- Wolfe, S., Ingold, C. F. & Lemieux, R. U. Oxidation of olefins by potassium permanganate: mechanism of α -ketol formation. *J. Am. Chem. Soc.* **103**, 938–939 (1981).
- Banoo, F. & Stewart, R. Mechanisms of permanganate oxidation. IX. Permanganate oxidation of aromatic alcohols in acid solution. *Can. J. Chem.* **47**, 3199–3205 (1969).
- Endo, M. Grow carbon fibers in the vapor phase. *Chemtech* **18**, 568–576 (1988).
- Hummers, W. S. Jr & Offeman, R. E. Preparation of graphitic oxide. *J. Am. Chem. Soc.* **80**, 1339 (1958).
- Lerf, A., He, H., Forster, M. & Klinowski, J. Structure of graphite oxide revisited. *J. Phys. Chem. B* **102**, 4477–4482 (1998).

25. Li, D., Mueller, M. B., Gilje, S., Kaner, R. B. & Wallace, G. G. Processable aqueous dispersions of graphene nanosheets. *Nature Nanotechnol.* **3**, 101–105 (2008).
26. Stankovich, S. *et al.* Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. *Carbon* **45**, 1558–1565 (2007).
27. Bourlino, A. B. *et al.* Graphite oxide: chemical reduction to graphite and surface modification with primary aliphatic amines and amino acids. *Langmuir* **19**, 6050–6055 (2003).
28. Eda, G., Fanchini, G. & Chhowalla, M. Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material. *Nature Nanotechnol.* **3**, 270–274 (2008).
29. Stankovich, S. *et al.* Stable aqueous dispersions of graphitic nanoplatelets via the reduction of exfoliated graphite oxide in the presence of poly(sodium 4-styrenesulfonate). *J. Mater. Chem.* **16**, 155–158 (2006).
30. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank P. M. Ajayan, W. Guo, J. Duque, Z. Sun, and Z. Jin for technical assistance and discussions. Mitsui & Co. generously donated the MWCNTs. The work was funded by the US Defense Advanced Research Projects Agency, the US Federal Aviation Administration, Department of Energy (DE-FC-36-05GO15073) and Wright Patterson Air Force Laboratory through the US Air Force Office of Scientific Research.

Author Contributions D.V.K. discovered the unzipping reaction and made most of the analysed ribbons. A.L.H. obtained and analysed most of the analysis data including the TEM, AFM, ultraviolet, XPS, TGA, Raman and infrared data; she also made some of the ribbons and wrote the majority of the manuscript. A.S. fabricated and tested the electronic devices. J.R.L. performed some of the spectral analysis including the X-ray diffraction. A.D. prepared and studied the nanoribbons on silicon surfaces for electrical analysis. B.K.P. performed some of the AFM analyses. J.M.T. oversaw and directed all aspects of the syntheses, data analysis and manuscript correction and finalization.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.T. (tour@rice.edu).

METHODS

MWCNTs were used as received from Mitsui & Co. (lot no. 05072001K28). The remaining chemicals were purchased from Sigma-Aldrich and used as received, except for the concentrated sulphuric acid (Fisher Scientific). Deionized water (18-M Ω resistivity) obtained from a NanoPure system (Barnstead) was used throughout this work.

Nanoribbon formation. In a typical procedure, MWCNTs (150 mg, 12.5 mequiv. carbon) were suspended in 150 ml of concentrated H₂SO₄ for 1 h to 12 h. KMnO₄ (750 mg, 4.75 mmol) was then added and the mixture allowed to stir for 1 h at room temperature. The reaction was then heated in an oil bath at 55 °C for 30 min. The progress of the reaction was monitored by preparing two test tubes, one containing 1 ml of deionized water and 2–3 drops of hydrogen peroxide (30%), and one containing 1 ml of water only. Four or five drops of the reaction mixture were added to the test tubes and bath-sonicated for 3 min. The colour and dispersibility of the solution in the test tube containing the hydrogen peroxide was noted; if the solution was yellow/brown and the solid broke up into very small pieces, the reaction was complete. The other test tube was used to check the level of permanganate consumption; if a dark purple/blue hue was observed (like that seen when KMnO₄ is dissolved in water), the reaction was not complete. Alternatively, a solution with a red hue signified the presence of manganese(IV), indicating permanganate consumption and a complete reaction. It was also important to note the colour of the reaction mixture itself; the colour went from black to dark brown with the green colour of permanganate in acid disappearing by reaction completion. If the reaction had not completed after 30 min at 55 °C, the temperature was increased to 65 °C. When the reaction temperature stabilized to 65 °C, the reaction status was checked again using the above test-tube procedure. Continued heating at 65 °C was needed if the permanganate was not entirely consumed.

When the reaction appeared nearly complete or complete, the temperature was increased to 70 °C and the solution was allowed to stabilize. Upon stabilization to 70 °C, the reaction was removed from the heat source, allowed to cool to room temperature and poured onto 400 ml of ice containing 5 ml of 30% H₂O₂ (to prevent precipitation of insoluble MnO₂). After vacuum filtration through a PTFE membrane (5.0- μ m pore size), the solid was removed and stirred in 150 ml of water for 30 min, and then bath-sonicated (Cole Parmer ultrasonic cleaner, Model 08849-00) for 15 min. The material was then flocculated by addition of 20 vol% concentrated HCl (30 ml). The solid was then filtered through a PTFE membrane (0.45- μ m pore size). The product was removed and stirred in 150 ml of ethanol for 30 min, then bath-sonicated (device as above) for 15 min. The material was flocculated by addition of 100 vol% ether (150 ml) followed by filtration through a PTFE membrane (0.45- μ m pore size). The final product was washed twice with ether (50 ml each time) and dried *in vacuo* to afford 321 mg of oxidized nanoribbons. Shortened ribbons (Fig. 1d) were formed from an aqueous solution of washed nanoribbons that were tip-sonicated at 30-W power (Misonix Sonicator 3000) for a total of 30 min (2 min on and 1 min off, for a total on-time of 30 min).

Stepwise oxidation. The stepwise oxidation was performed by first running the reaction, under the conditions specified above, with a 1:1 mass ratio of KMnO₄:MWCNT (denoted sample I, least oxidized). After 1 h at 55 °C, approximately one-fifth of the volume of the reaction mixture was extracted and worked up as outlined above. The remaining portion was then treated with 100 wt% KMnO₄ (relative to the portion of MWCNTs remaining) and allowed to react for

another 1 h at 55 °C. After extracting one-quarter of the volume of the reaction mixture, to be worked up, the rest was treated again with 100 wt% KMnO₄. The process was repeated until the final batch of ribbons was exposed to a total of 500 wt% KMnO₄ (denoted sample V, most oxidized; Fig. 2e).

Reduction of nanoribbons. Reduction in aqueous N₂H₄ can be done with or without SDS surfactant. To introduce the SDS surfactant, 50 mg of the as-prepared nanoribbons were homogenized (IKA T-25 digital ULTRA-TURRAX disperser with 18 G dispersing element, 7,000 r.p.m.) in 250 ml of 1% aqueous SDS solution for 1 h. The suspension was then cup-sonicated (Cole Parmer ultrasonic processor, 75% power) for 10 min and filtered through a 5-cm plug of glass wool. If SDS was not present, 50 mg of as-prepared nanoribbons was dissolved in 250 ml nanopure water. In both cases, 150 ml of the nanoribbon solution was covered by a 5-mm layer of silicon oil in an Erlenmeyer flask. To this solution we added 150 μ l of concentrated NH₄OH (2.22 mmol), followed by 150 μ l of N₂H₄·H₂O (98%, 3.03 mmol). The oil was added to eliminate the air–water interface that causes agglomeration of the reduced ribbons as water evaporates²⁵. The reaction mixture was heated at 95 °C for 1 h in a gently boiling water bath (no stirring). Excess N₂H₄ was removed by dialysis (CelluSep H1 regenerated cellulose tubular membrane; nominal molecular weight, 5,000 (relative molecular mass); 400-mm width) for 24 h in 1% NH₄OH. Filtering and drying *in vacuo* afforded 23.2 mg of reduced nanoribbons. Therefore, the overall weight conversion efficiencies from MWCNTs to the reduced nanoribbons were 215% and 46% for the two steps, yielding 99%.

Device fabrication. Fabrication of graphene devices was performed by tracking individual nanoribbons on the surface of highly doped Si substrates, covered with 200-nm-thick dielectric SiO₂, by SEM (JEOL-6500 microscope), followed by patterning of 20-nm-thick Pt contacts by standard electron beam lithography. Before being tested, the devices used to produce the data in Fig. 4d were annealed in H₂/Ar (1:1, <1 atm) atmosphere at 300 °C for 10 min. The electrical properties were tested using a probe station (Desert Cryogenics TT-probe 6 system) under vacuum with a chamber base pressure below 10^{−5} mm Hg. The current–voltage data were collected using an Agilent 4155C semiconductor parameter analyser.

Sample analysis. TEM imaging was performed on a JEOL-2010 microscope. Samples were prepared by dispersing the nanoribbons in a 1:1 ethanol:water mixture and dropped onto 300 mesh holey, lacey carbon grids on copper support (Ted Pella). AFM images were obtained with a Nanoscope IIIa (Digital Instruments/Veeco Metrology), operating in tapping mode, using Si tips n-doped with 1–10 Ω cm phosphorus (Veeco, MPP-11100-140) at a scan rate of 2 Hz and a resolution of 512 \times 512. Samples for AFM analysis were prepared by spin-coating aqueous solutions of nanoribbons at 3,000 r.p.m. onto a freshly cleaved mica surface (Ted Pella) and rinsed with deionized water and 2-isopropanol.

Ultraviolet–visible spectra were obtained on a Shimadzu UV-3101 PC with samples contained in 1-ml quartz cuvettes. Fourier-transform infrared spectroscopy was performed using a Nicolet FTIR Infrared Microscope with an ATR attachment. XPS was performed on a PHI Quantera SXM scanning X-ray microprobe with a pass energy of 26.00 eV, 45° take-off angle and a 100- μ m beam size. TGA (Q50, TA Instruments) was performed from room temperature to 950 °C at 10 °C min^{−1} under argon. Raman spectroscopy was performed on a Renishaw Raman scope using a 633-nm HeNe laser. The X-ray diffraction measurements were carried out with a Rigaku diffractometer (Cu K α radiation; X-ray wavelength, λ = 1.5406 Å; operating energy, 40 keV; cathode current, 40 mA; scan rate, 1° min^{−1}). Brunauer–Emmett–Teller surface-area analysis was performed at 77 K on a Quantachrome Autosorb-1 Physisorption system using N₂ as the sorption gas.

Narrow graphene nanoribbons from carbon nanotubes

Liyang Jiao^{1*}, Li Zhang^{1*}, Xinran Wang¹, Georgi Diankov¹ & Hongjie Dai¹

Graphene nanoribbons (GNRs) are materials with properties distinct from those of other carbon allotropes^{1–5}. The all-semiconducting nature of sub-10-nm GNRs could bypass the problem of the extreme chirality dependence of the metal or semiconductor nature of carbon nanotubes (CNTs) in future electronics^{1,2}. Currently, making GNRs using lithographic^{3,4,6}, chemical^{7–9} or sonochemical¹ methods is challenging. It is difficult to obtain GNRs with smooth edges and controllable widths at high yields. Here we show an approach to making GNRs by unzipping multi-walled carbon nanotubes by plasma etching of nanotubes partly embedded in a polymer film. The GNRs have smooth edges and a narrow width distribution (10–20 nm). Raman spectroscopy and electrical transport measurements reveal the high quality of the GNRs. Unzipping CNTs with well-defined structures in an array will allow the production of GNRs with controlled widths, edge structures, placement and alignment in a scalable fashion for device integration.

The high carrier mobility of graphene^{10–14} offers the possibility of building high-performance graphene-based electronics. Recently, both theoretical^{15–18} and experimental^{1–4} works have shown that quantum confinement and edge effects introduce a band gap in narrow graphene ribbons independent of chirality, and the resulting GNR semiconductors can be used to make field-effect transistors. Several approaches have been developed to obtain GNRs. Lithographic patterning has been used to produce wide ribbons (>20 nm) from graphene sheets^{3,4}, but the width and smoothness of the GNRs were limited by the resolution of the lithography and etching techniques. Bulk amounts of wide (20–300 nm) and few-layered (2–40) GNRs were synthesized by a chemical vapour deposition method⁹. A chemical sonication route developed by our group produced sub-10-nm GNR semiconductors from intercalated and exfoliated graphite¹. However, the yield of GNRs was low and their width distribution was broad; widths ranged from less than 10 nm to ~100 nm.

Because CNTs are considered to be GNRs rolled up into seamless tubes and the synthesis, size control, placement and alignment control of nanotubes have been widely investigated and established^{19,20}, we ask the question of whether CNTs can be unzipped to form GNRs with structural control. The greatest challenge in converting CNTs to GNRs is to develop ways of cleaving CNTs in the longitudinal direction without rapid etching along the circumference.

Here we develop controlled unzipping of CNTs to produce GNRs by an Ar plasma etching method. To remove a longitudinal strip of carbon atoms from the side walls of CNTs, we first embedded multi-walled carbon nanotubes (MWCNTs) in a poly(methyl methacrylate) (PMMA) layer as an etching mask (Fig. 1). Briefly, pristine MWCNTs (diameter, ~4–18 nm; see Supplementary Fig. 1a) synthesized by arc discharge (Bucky tube, Aldrich) were dispersed in 1% surfactant solution by brief sonication and deposited onto a Si

substrate. A 300-nm-thick film of PMMA was spin-coated on top of the MWCNTs. After baking, the PMMA–MWCNT film was peeled off in a KOH solution²¹ (Fig. 1b). MWCNTs embedded in the resulting PMMA film had a narrow strip of side wall not covered by PMMA, owing to conformal PMMA coating on the substrate. The PMMA–MWCNT film was then exposed to a 10-W Ar plasma for various times (Fig. 1c). Owing to protection by the PMMA, the top side walls of MWCNTs were etched faster and removed by the plasma. Single-, bi- and multilayer GNRs and GNRs with inner CNT cores (Fig. 1d–g) were produced depending on the diameter and number of layers of the starting MWCNT and the etching time. After etching, the PMMA film was contact-printed and attached to a Si substrate with a 500-nm-thick layer of SiO₂. Finally, the PMMA film was removed using acetone vapour²¹, and this was followed by calcination at 300 °C for 10 min to remove polymer residue on the target substrate, leaving GNRs on the substrate (Fig. 1h).

The mean diameter of our starting MWCNTs was ~8 nm (Fig. 2a and Supplementary Fig. 1a). For 10 s of Ar plasma etching, our method converted ~20% of the MWCNTs into single- or few-layer GNRs of 10–20-nm width (height, <2 nm; see Fig. 2). Other products were multilayer GNRs or GNRs with CNT cores (Fig. 1d). The yield of GNRs was limited by the relatively wide diameter distribution of the starting MWCNTs used (Supplementary Information). All of the width data were obtained after correcting the tip-size effect¹. Within

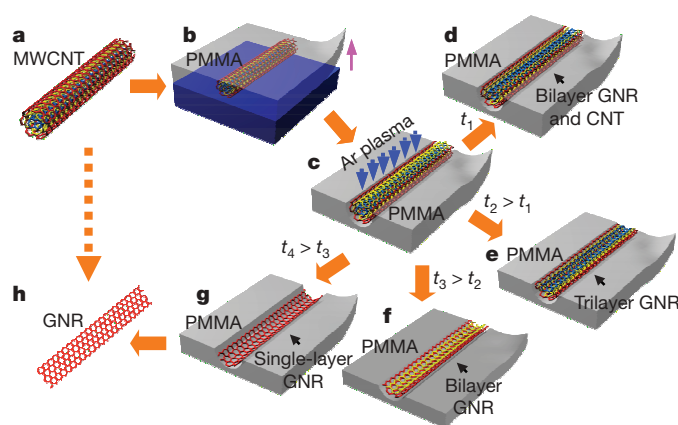


Figure 1 | Making GNRs from CNTs. **a**, A pristine MWCNT was used as the starting raw material. **b**, The MWCNT was deposited on a Si substrate and then coated with a PMMA film. **c**, The PMMA–MWCNT film was peeled from the Si substrate, turned over and then exposed to an Ar plasma. **d–g**, Several possible products were generated after etching for different times: GNRs with CNT cores were obtained after etching for a short time t_1 (**d**); tri-, bi- and single-layer GNRs were produced after etching for times t_2 , t_3 and t_4 , respectively ($t_4 > t_3 > t_2 > t_1$; **e–g**). **h**, The PMMA was removed to release the GNR.

¹Department of Chemistry and Laboratory for Advanced Materials, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

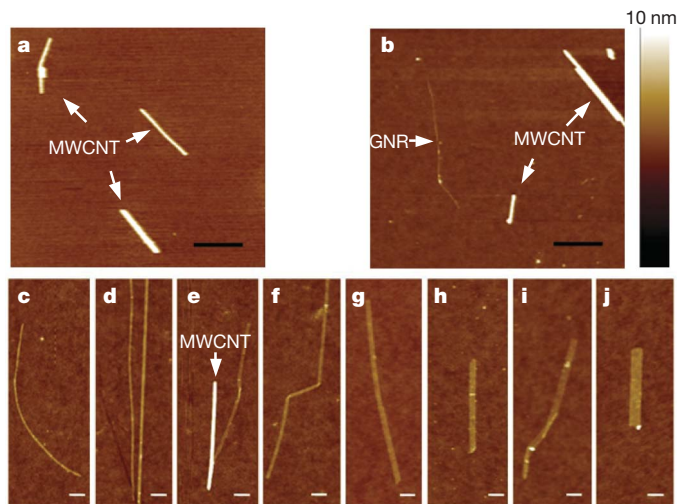


Figure 2 | Images of GNRs converted from MWCNTs. **a**, An AFM image of raw MWCNTs dispersed on a Si substrate. **b**, An image of the substrate after the GNR conversion process, showing coexistence of MWCNTs and GNRs. Scale bars, 1 μm . **c–j**, Single- or few-layer GNRs of different widths and heights: respectively 7 and 1.8 nm (**c**), 8 and 1.8 nm (left, **d**), 13 and 2.0 nm (right, **d**), 15 and 0.9 nm (**e**), 17 and 1.0 nm (**f**), 25 and 1.1 nm (**g**), 33 and 1.4 nm (**h**), 45 and 0.8 nm (**i**) and 51 and 1.9 nm (**j**). Scale bars, 100 nm. The height scale for all the AFM images is 10 nm. In **e**, an arrow points to a leftover MWCNT.

the resolution of atomic force microscopy (AFM), the edges of the obtained GNRs were very smooth. Unlike for previous GNRs, the ribbons were uniform in width along their lengths, none being wedge shaped¹, owing to the quasi-one-dimensional CNT templates with uniform diameter along their lengths. According to histogram data on GNRs (Supplementary Fig. 1b, c), the widths of our GNRs mostly fell into the range 10–20 nm, which is narrower than that of GNRs made by the sonochemical method¹. The width of 10–20 nm corresponds to half of the circumference of the starting MWCNTs with the

mean diameter of ~ 8 nm. The heights of our GNRs were ~ 0.8 , 1.3 and 1.8 nm, which we assign to single-, bi- and trilayer GNRs, respectively²². The folding observed in single-layer GNRs (Fig. 2e, f, i) suggests that the flexibility of the GNRs is excellent in comparison with that of MWCNTs. The folds resulted from bending of the PMMA film²³ and/or the perturbation of liquid during the removal of the PMMA in acetone.

Raman spectroscopy is a powerful approach to investigating the structural and electronic properties of carbon-based materials. The shape of the second-order Raman band (2D) is a characteristic that can be used to distinguish the number of layers of AB-stacked graphene^{24–26}. We carried out confocal Raman mapping on single-, bi- and trilayer GNRs (Methods). We obtained an AFM image (Fig. 3a) and a G-band image (Fig. 3b) of a 0.9-nm-thick GNR on the same length scale. The 2D peak was well fitted by a sharp and symmetric Lorentzian peak (inset, Fig. 3c), suggesting the single-layer nature of the GNR. For another 1.3-nm-thick GNR (Fig. 3d–f), the 2D peak was fitted by four Lorentzians (inset, Fig. 3f), characteristic of bilayer, AB-stacked graphene. Some bilayer GNRs showed 2D peaks with different line shapes from AB-stacked graphene, reflecting the varying stacking structures between layers in the starting MWCNTs (Supplementary Fig. 2). For a trilayer GNR (height, ~ 1.9 nm; see Supplementary Fig. 3), similar to trilayer, ABA-stacked graphene sheets, the 2D band was broader and up shifted as the thickness increased (Supplementary Table 1). We note that the 2D band of the MWCNTs was broader than that of our GNRs (Supplementary Fig. 6), as expected.

The intensity ratio of the D and G bands (I_D/I_G) is commonly used to evaluate the quality of carbon materials. No obvious D band was observed for any of the MWCNTs dispersed on substrates (Supplementary Fig. 6), suggesting the high quality of the starting materials. The average I_D/I_G values were respectively 0.38, 0.30 and 0.28 for our single-, bi- and trilayer GNRs with 10–20-nm widths. Because defect density on the pristine MWCNTs was low, the D-band Raman signal of our GNRs should be mainly due to their open edges^{24,25}. The I_D/I_G values were lower than those for GNRs obtained

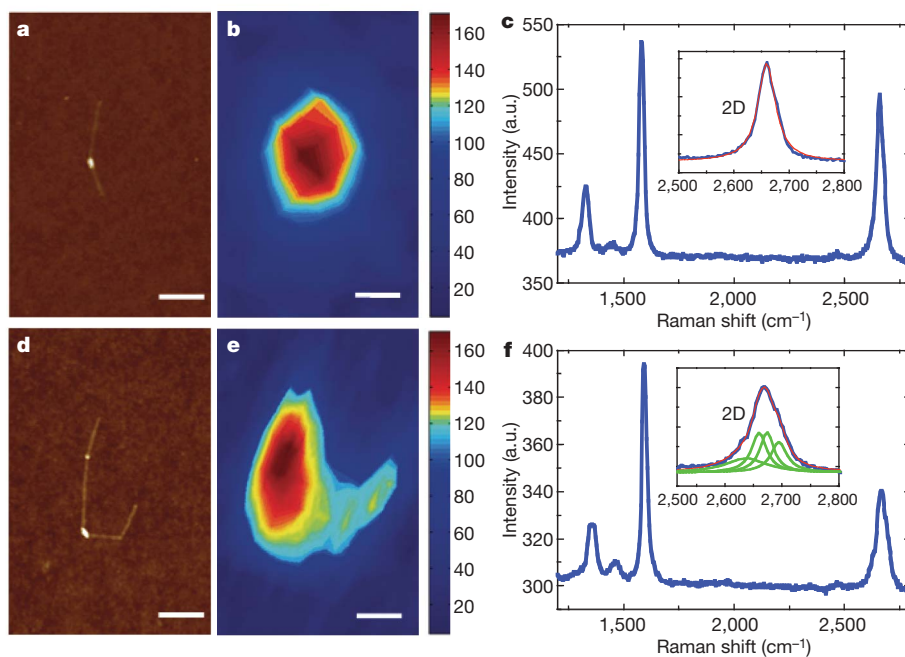


Figure 3 | Raman imaging and spectra of GNRs. **a–c**, An AFM image (**a**), a G-band Raman image (**b**) and the Raman spectrum (**c**) of a single-layer GNR. Inset of **c**, a 2D-band spectrum (blue) of the GNR and a single-Lorentzian fit (red). a.u., arbitrary units. **d–f**, An AFM image (**d**), a G-band Raman image (**e**) and the Raman spectrum (**f**) of a bilayer GNR. Inset of **f**, 2D band

spectrum (blue) of the GNR and the best-fit curve (red), which is a combination of four Lorentzians (green). The $\sim 1,440\text{ cm}^{-1}$ peaks in **c** and **f** are attributed to trace PMMA left on the SiO_2 (Supplementary Fig. 4). Scale bars, 200 nm.

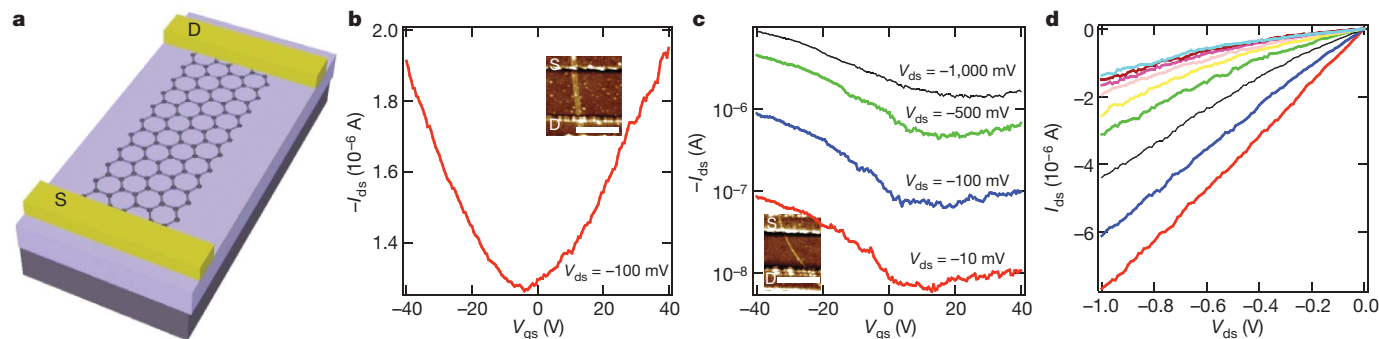


Figure 4 | Room-temperature electrical properties of GNR devices. **a**, A GNR device. S, source; D, drain. **b**, Plot of drain–source current (I_{ds}) versus gate–source voltage (V_{gs}) for a ~ 16 -nm-wide GNR device probed in vacuum after electrical annealing. The Dirac point is near $V_g = 0$ V. Inset, AFM image of this device; scale bar, 200 nm. **c**, I_{ds} – V_{gs} curves for a ~ 7 -nm-wide GNR

device at various biases probed in air (inset, AFM image; scale bar, 200 nm). The ratio of on-state current (I_{on}) to off-state current (I_{off}) for this GNR device is ~ 10 . **d**, I_{ds} – V_{ds} curves for the device in **c** at gate biases V_{gs} ranging from -40 V (bottom) to 40 V (top) in steps of 10 V.

by lithographic etching of pristine graphene sheets (I_D/I_G was ~ 2 for a bilayer GNR of ~ 28 -nm width obtained by lithographic patterning (X. R. Wang *et al.*, unpublished data); see Supplementary Fig. 5), indicating the high quality of our GNRs, in particular the high degree of edge smoothness.

There are several key steps in making GNRs from MWCNTs using our approach. Embedding MWCNTs in a PMMA film as an etching mask is essential to protect parts of the shells of the MWCNTs from plasma. The relatively low viscosity and good wetting capability of PMMA allowed its conformal coating on the MWCNTs. The choice of plasma source is also critical to longitudinally unzip CNTs. O_2 plasma was widely used for chemical etching of carbon materials. However, control experiments found that O_2 plasma etching tended to remove complete shells from MWCNTs, even with the partial protection of PMMA (Supplementary Fig. 7), suggesting that the chemical etching of MWCNTs by O_2 plasma was rapid along the nanotube circumference and could not be used to produce GNRs. The bombardment effect of Ar plasma²⁷ offered anisotropic physical etching to remove atoms at unprotected sites along the longitudinal direction of the CNTs. To make the etching more controllable, we used a low plasma power of 10 W and optimized the etching time. We found that longer etching duration increased the yield of single- and few-layer GNRs, but caused breaks in the GNRs. After etched for 20 s and longer, ribbons tended to be discontinuous and have frequent cuts. Raman spectra of GNRs obtained by etching for 30 s showed high I_D/I_G values, of >1.0 (Supplementary Fig. 8). Inhomogeneity in MWCNT diameter and number of layers caused variation in the GNR number of layers and the residue CNT structures. By using a starting CNT material with a controlled shell number and diameter, we expect to obtain GNRs of a well-defined width and number of layers at $\sim 100\%$ yield.

We fabricated three-terminal devices with our GNRs (Fig. 4), with Pd as source and drain contacts (channel length, $L \approx 250$ nm), a p++ Si back gate and thermally grown 500 -nm SiO_2 as gate dielectrics. GNRs ≤ 10 nm in width showed field-effect transistor characteristics with p-doping effects due to the physisorbed O_2 from ambient and other species during the treatment steps (Fig. 4c and Supplementary Fig. 9). A ~ 7 -nm-wide GNR device had an I_{on}/I_{off} ratio of >10 (Fig. 4c) and a ~ 6 -nm-wide GNR device had $I_{on}/I_{off} > 100$ (Supplementary Fig. 9). These GNRs had quantum-confined semiconducting characteristics, rather than those of bulk graphene, with much weaker gate modulation of conductance. We note that the error bar in widths measured by AFM was ± 2 – 3 nm, which led to relatively large uncertainties in the measured widths in the sub- 10 -nm region. To obtain GNR devices with higher I_{on}/I_{off} values, even narrower GNRs are needed and could be obtained by our method by using few-walled CNTs.

Our >10 -nm-wide GNR devices showed weaker gate dependence (Fig. 4b) owing to smaller band gaps, consistent with lithographically

and chemically derived GNRs of similar widths^{1,3,4}. Figure 4b shows the transfer characteristics of a ~ 16 -nm-wide GNR device probed in vacuum after electrical annealing²⁸ by applying a bias of up to 3 V. The GNR exhibited a clear conductance minimum corresponding to the Dirac point at $V_g \approx 0$ V after electrical annealing by removing physisorbed O_2 and other p-doping species (Supplementary Fig. 10). The ‘V’-shaped I_{ds} – V_{gs} curve of the GNR device resembles that of bulk graphene, reflecting symmetric hole and electron transports at negative and positive gate voltages, respectively. The resistivity of the devices at the Dirac point for our 10 – 20 -nm-wide GNRs was 10 – 40 k Ω , similar to lithographically patterned 20 – 50 -nm-wide GNRs^{3,22,29}. The fact that the resistivity of our GNRs is comparable to that of the lithographically patterned ones suggests a similar quality of GNRs in the 10 – 20 -nm width range. The mobilities of the GNRs made by both methods were also similar and ~ 10 times lower than those of large two-dimensional graphene sheets, most likely because of edge scattering in the GNRs^{10–12}.

Unzipping of CNTs offers a new way of producing GNRs with controlled structure and quality. Our approach is also compatible with semiconductor processing. Progress made in the synthesis, size control, placement and alignment control of CNTs can be exploited to make GNRs in a controlled fashion. CNTs with narrower diameter and chirality distributions can be used to make GNRs with well-defined widths and edge structures. Few-walled CNTs can be used to obtain narrow, sub- 10 -nm, GNRs with band gaps sufficient for room-temperature transistor applications. Aligned CNT arrays can lead to GNR arrays. Thus, it should be possible to produce large-scale, well-aligned semiconducting GNRs with controlled structures for practical application in electronics.

METHODS SUMMARY

Preparation of GNRs. We dispersed 1 mg of MWCNTs in 10 ml of 1% Tween 20 aqueous solution by sonication for 5 min and then centrifugation at $16,400g$ for 10 min to remove aggregates. The MWCNT suspension was deposited onto a Si substrate pretreated with 3-aminopropyltriethoxysilane (APTES, 12 ml in 20 ml of H_2O), rinsed with water and then blow-dried. The sample was then calcined at $350^\circ C$ for 10 min to remove the Tween 20. A PMMA solution (relative molecular mass, $M_w = 495,000$; 5% in anisole) was spin-coated on MWCNTs on the substrate at $3,000$ rounds per minute for 1 min and then baked at $170^\circ C$ for 2 h on a hot plate. The PMMA–MWCNT film was peeled off in 1 M KOH solution at $80^\circ C$ (ref. 21). Then the film was rinsed with water and printed onto a Si substrate. To make the film well-adhered to the Si substrate, we heated the sample at $80^\circ C$ for 10 min in an oven. Ten-watt Ar plasma was used to etch the PMMA–MWCNT film at the base pressure of 40 mTorr. After etching, we lifted the PMMA film using water and then adhered it to an APTES-treated, 500 -nm SiO_2 /Si substrate with a prefabricated Pt/W marker array. After the PMMA had been removed using acetone vapour²¹, the obtained sample was calcined at $300^\circ C$ for 10 min to remove the residue of PMMA.

Fabrication of GNR devices. We located the GNRs with AFM and recorded their locations relative to the prefabricated Pt/W markers. The source–drain electrode

pattern was then designed to provide electrical contact to the GNRs; this was carried out by electron beam lithography, 20-nm Pd metal deposition and lift off. The devices were then annealed in Ar at 220 °C for 15 min to improve the contact quality. Electrical characterization of the devices was carried out both in air and in vacuum using a semiconductor analyser (Agilent 4156C).

Characterization of GNRs using Raman spectroscopy. Raman spectra of individual GNRs located by AFM were measured using a 633-nm HeNe laser (spot size, $\sim 1\ \mu\text{m}$) and a power of $\sim 1\ \text{mW}$. The mapping area was set to be $3\ \mu\text{m} \times 3\ \mu\text{m}$ and the step size was 100 nm. The D, G and 2D bands were recorded and the integration time was 10 s at each spot.

Received 30 December 2008; accepted 16 February 2009.

- Li, X. L. *et al.* Chemically derived, ultrasmooth graphene nanoribbon semiconductors. *Science* **319**, 1229–1232 (2008).
- Wang, X. R. *et al.* Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors. *Phys. Rev. Lett.* **100**, 206803 (2008).
- Chen, Z. H., Lin, Y. M., Rooks, M. J. & Avouris, P. Graphene nano-ribbon electronics. *Physica E (Amsterdam)* **40**, 228–232 (2007).
- Han, M. Y., Ozyilmaz, B., Zhang, Y. B. & Kim, P. Energy band-gap engineering of graphene nanoribbons. *Phys. Rev. Lett.* **98**, 206805 (2007).
- Cresti, A. *et al.* Charge transport in disordered graphene-based low dimensional materials. *Nano Res.* **1**, 361–394 (2008).
- Tapasztó, L., Dobrik, G., Lambin, P. & Biro, L. P. Tailoring the atomic structure of graphene nanoribbons by scanning tunnelling microscope lithography. *Nature Nanotechnol.* **3**, 397–401 (2008).
- Datta, S. S., Strachan, D. R., Khamis, S. M. & Johnson, A. T. C. Crystallographic etching of few-layer graphene. *Nano Lett.* **8**, 1912–1915 (2008).
- Ci, L. J. *et al.* Controlled nanocutting of graphene. *Nano Res.* **1**, 116–122 (2008).
- Campos-Delgado, J. *et al.* Bulk production of a new form of sp^2 carbon: crystalline graphene nanoribbons. *Nano Lett.* **8**, 2773–2778 (2008).
- Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Zhang, Y. B., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Berger, C. *et al.* Electronic confinement and coherence in patterned epitaxial graphene. *Science* **312**, 1191–1196 (2006).
- Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6**, 183–191 (2007).
- Nakada, K., Fujita, M., Dresselhaus, G. & Dresselhaus, M. S. Edge state in graphene ribbons: nanometer size effect and edge shape dependence. *Phys. Rev. B* **54**, 17954–17961 (1996).
- Barone, V., Hod, O. & Scuseria, G. E. Electronic structure and stability of semiconducting graphene nanoribbons. *Nano Lett.* **6**, 2748–2754 (2006).
- Son, Y. W., Cohen, M. L. & Louie, S. G. Energy gaps in graphene nanoribbons. *Phys. Rev. Lett.* **97**, 216803 (2006).
- Yang, L. *et al.* Quasiparticle energies and band gaps in graphene nanoribbons. *Phys. Rev. Lett.* **99**, 186801 (2007).
- Dai, H. J. Carbon nanotubes: opportunities and challenges. *Surf. Sci.* **500**, 218–241 (2002).
- Jorio, A., Dresselhaus, M. S. & Dresselhaus, G. *Carbon Nanotubes: Advanced Topics in the Synthesis, Structure, Properties and Applications*. (Springer, 2008).
- Jiao, L. Y. *et al.* Creation of nanostructures with poly(methyl methacrylate)-mediated nanotransfer printing. *J. Am. Chem. Soc.* **130**, 12612–12613 (2008).
- Lin, Y. M. & Avouris, P. Strong suppression of electrical noise in bilayer graphene nanodevices. *Nano Lett.* **8**, 2119–2125 (2008).
- Jiao, L. Y., Xian, X. J. & Liu, Z. F. Manipulation of ultralong single-walled carbon nanotubes at macroscale. *J. Phys. Chem. C* **112**, 9963–9965 (2008).
- Ferrari, A. C. *et al.* Raman spectrum of graphene and graphene layers. *Phys. Rev. Lett.* **97**, 187401 (2006).
- Graf, D. *et al.* Spatially resolved Raman spectroscopy of single- and few-layer graphene. *Nano Lett.* **7**, 238–242 (2007).
- Ni, Z. H., Wang, Y. Y., Yu, T. & Shen, Z. X. Raman spectroscopy and imaging of graphene. *Nano Res.* **1**, 273–291 (2008).
- Winters, H. F., Coburn, J. W. & Chuang, T. J. Surface processes in plasma-assisted etching environments. *J. Vac. Sci. Technol. B* **1**, 469–480 (1983).
- Moser, J., Barreiro, A. & Bachtold, A. Current-induced cleaning of graphene. *Appl. Phys. Lett.* **91**, 163513 (2007).
- Lin, Y. M., Perebeinos, V., Chen, Z. H. & Avouris, P. Electrical observation of subband formation in graphene nanoribbons. *Phys. Rev. B* **78**, 161409 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Microelectronics Advanced Research Corporation - Materials, Structures, and Devices Center, Intel and the US Office of Naval Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.D. (hdai@stanford.edu).

Rapid sea-level rise and reef back-stepping at the close of the last interglacial highstand

Paul Blanchon¹, Anton Eisenhauer², Jan Fietzke² & Volker Liebetrau²

Widespread evidence of a +4–6-m sea-level highstand during the last interglacial period (Marine Isotope Stage 5e) has led to warnings that modern ice sheets will deteriorate owing to global warming and initiate a rise of similar magnitude by AD 2100 (ref. 1). The rate of this projected rise is based on ice-sheet melting simulations and downplays discoveries of more rapid ice loss^{2,3}. Knowing the rate at which sea level reached its highstand during the last interglacial period is fundamental in assessing if such rapid ice-loss processes could lead to future catastrophic sea-level rise. The best direct record of sea level during this highstand comes from well-dated fossil reefs in stable areas^{4–6}. However, this record lacks both reef-crest development up to the full highstand elevation, as inferred⁷ from widespread intertidal indicators at +6 m, and a detailed chronology, owing to the difficulty of replicating U-series ages on submillennial timescales⁸. Here we present a complete reef-crest sequence for the last interglacial highstand and its U-series chronology from the stable northeast Yucatán peninsula, Mexico. We find that reef development during the highstand was punctuated by reef-crest demise at +3 m and back-stepping to +6 m. The abrupt demise of the lower-reef crest, but continuous accretion between the lower-lagoonal unit and the upper-reef crest, allows us to infer that this back-stepping occurred on an ecological timescale and was triggered by a 2–3-m jump in sea level. Using strictly reliable ²³⁰Th ages of corals from the upper-reef crest, and improved stratigraphic screening of coral ages from other stable sites, we constrain this jump to have occurred ~121 kyr ago and conclude that it supports an episode of ice-sheet instability during the terminal phase of the last interglacial period.

Reconstruction of eustatic sea level from the precise age and elevation of submerged reef-crest corals has shown that the last deglaciation (Termination 1) was punctuated by rapid, metre-scale rise events caused by ice-sheet instability^{9–12}. Little is known, however, about the potential for unstable ice-sheet and sea-level behaviour within fully interglacial climates. The Holocene interglacial has been relatively stable, but the picture of sea-level variation during the last interglacial is less clear, owing to greater chronological uncertainty⁸. There is circumstantial evidence that it was punctuated by instability both during the preceding deglaciation¹³ and, more unexpectedly, during the highstand⁷. For example, highstand reef sequences in uplifted terraces show double-reef architectures, marine erosion surfaces and coral age–depth relationships suggestive of significant relative sea-level excursions^{14–16}. However, coseismic displacement and intraterrace faulting cannot be discounted as a cause of relative sea-level variability in such active terranes. Sea-level instability has also been inferred from an isotopic reconstruction from the Red Sea¹⁷, but these data have large uncertainties and are inconsistent with coral-based sea-level data from stable areas (Supplementary Discussion).

In these stable areas, well-dated reef terraces are developed to +3 m (refs 4–6), but intertidal deposits and notches indicate that

sea level reached +5–6 m during the highstand⁷. This absence of reefs above +3 m, along with the presence of similar erosion surfaces, has also been interpreted as evidence of sea-level excursions both during and terminating the highstand^{7,18}. Without a complete reefal sequence to +6 m, however, a convincing case for sea-level instability cannot be made. Pervasive open-system diagenesis makes it unlikely that the duration or rate of submillennial sea-level events can be radiometrically constrained during the last interglacial, as they are for Termination 1, but rapid accretion combined with a predictable depth zonation of coral and encruster assemblages¹⁹ means that reefs respond to and preserve extrinsic changes on ecological (decadal) timescales²⁰, making them ideal for identifying rapid changes in eustatic sea level associated with ice-sheet instability¹⁰.

To investigate possible submillennial changes in sea level during the last interglacial highstand, we analysed the stratigraphic architecture, palaeoecologic zonation and ²³⁰Th age structure of an exceptionally well-exposed fossil reef at Xcaret, a theme park on the northeast coast of the Yucatán peninsula. The northern part of the peninsula is an ideal location at which to investigate sea-level behaviour because it is historically aseismic²¹ and lacks neotectonic activity²², as confirmed by both its thick, surficial off-lapping sequence of undeformed Miocene to Pleistocene carbonates and its +6-m last interglacial shoreline²³. In addition, the conformity of this highstand elevation with other stable sites across the Caribbean⁷ indicates that the glacio-isostatic state of the peninsula, relative to an equilibrium state, was similar to the present²⁴ and that submillennial changes in relative sea-level history should accurately reflect a eustatic signature.

We measured 40 vertical sections and examined ~2.5 km of lateral exposure through the fossil reef (Fig. 1). For the first time in a stable area, these sections show a complete reefal sequence, consisting of two separate linear reef tracts with reef crests that are offset and at different elevations (Supplementary Fig. 1). The lower-reef tract crops out along northern shore for ~500 m and its reef crest extends from below mean sea level to +3 m. Like modern reefs¹⁹, the crest consists of large *A. palmata* colonies dispersed within an *A. palmata* boulder gravel; this is flanked by a reef front with a mixed coral assemblage and a large, lagoonal patch-reef complex. The upper-reef tract crops out ~150 m inland and parallels the southern section of shore for ~400 m. The reef crest, which also consists of large *A. palmata* colonies and its boulder-sized clasts, is developed between +3 and +5.8 m and is founded directly upon the lagoonal patch-reef complex of the lower tract. Both reef tracts are overlain by a regressive beach unit that prograded down slope as sea level fell from the highstand.

To distinguish chronologically between these two stages of reef-tract development, we obtained ²³⁰Th ages in triplicate from ten well-preserved coral samples, five from each tract (Supplementary Table 1). From the lower reef, two subsamples (xA4-2b and xA4-2c) gave isotopically reliable ages (Methods). By contrast, all five samples

¹Institute of Marine & Limnological Sciences, National Autonomous University of Mexico, AP1152, Cancun, 77500 Quintana Roo, Mexico. ²Leibniz Institute of Marine Science, IFM-GEOMAR, Wischhofstrasse 1–3, 24148 Kiel, Germany.

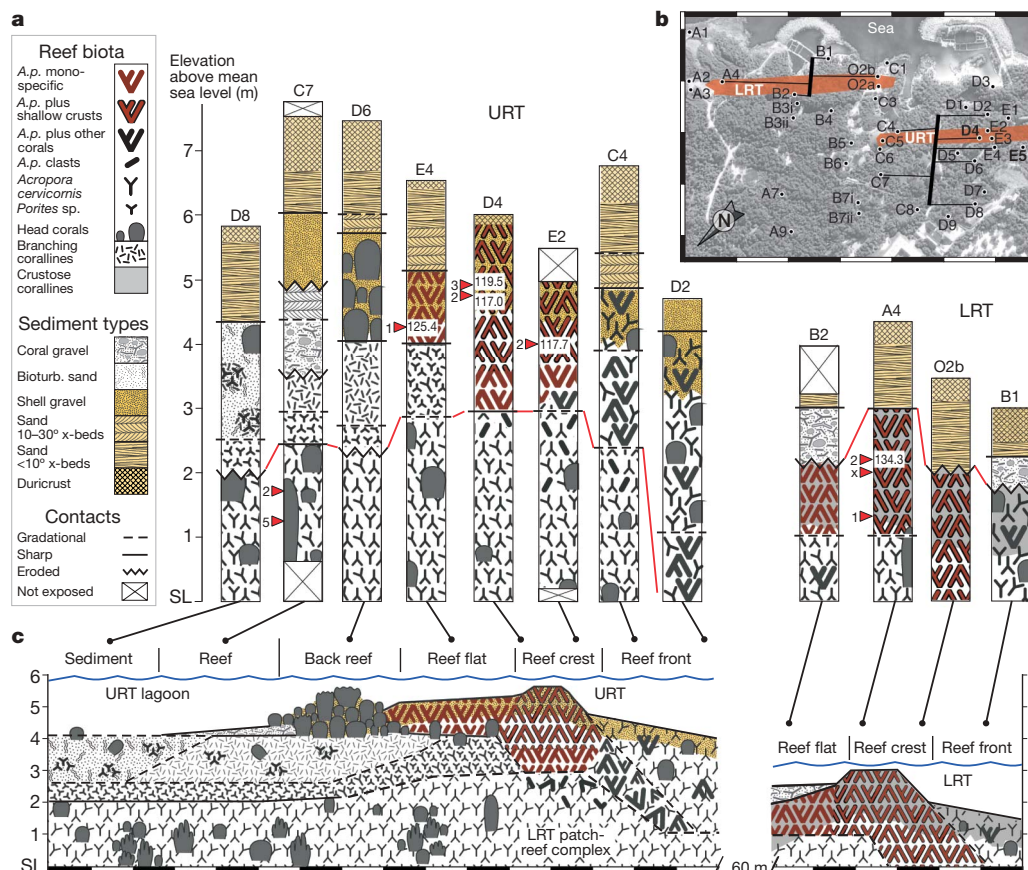


Figure 1 | Stratigraphic sections and reconstruction of reef development at Xcaret during the last interglacial. **a**, Stratigraphic sections along transects crossing *Acropora palmata* (*A.p.*) reef crests, showing age, geometry, elevation and contact relations of the upper-reef tract (URT) and the lower-reef tract (LRT) (red line marks boundary). Positions of dated coral samples (ages in kiloyears) are indicated by red arrowheads (only strictly reliable ages shown).

from the upper reef gave at least one age that is isotopically reliable (Fig. 2). These reliable ^{230}Th ages confirm that both reef tracts are contemporaneous and formed during Marine Isotope Stage 5e. However, the reliable subsamples from upper-reef corals reveal a millennial-scale age variability that exceeds the analytical error by a factor of

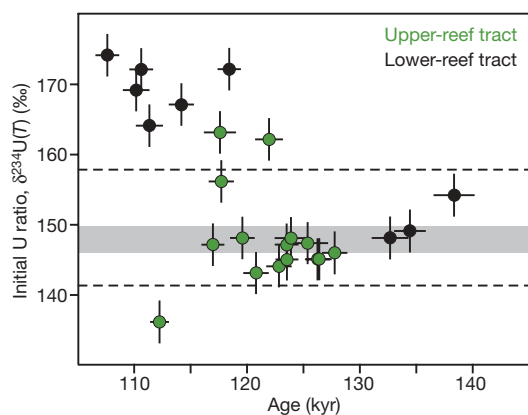


Figure 2 | Isotopic reliability of ^{230}Th coral ages from Xcaret. Ages within the grey band ($146.6\text{--}149.6\text{‰}$) have the same $\delta^{234}\text{U}(T)$ value as modern corals and sea water and are strictly reliable. $\delta^{234}\text{U}(T)$ represents the decay-corrected activity ratio calculated from the value measured today ($T = 0$): $\delta^{234}\text{U}(T) = \delta^{234}\text{U}(0)\exp(\lambda_{234}T)$, where $\delta^{234}\text{U}(0) (\text{‰}) = [(^{234}\text{U}/^{238}\text{U})_{\text{activity}} - 1] \times 10^3$ and $\lambda_{234} = 2.8263 \times 10^{-6} \text{ yr}^{-1}$. Ages within the dashed lines correspond to $149 \pm 8\text{‰}$ and are isotopically reliable to within 2 kyr (Methods). All other ages are unreliable. Error bars, two standard deviations of the mean (2σ).

3–10—a finding that is consistent with other analyses of true-age variability⁸. To address this variability, we excluded subsamples with $\delta^{234}\text{U}(T)$ values outside the range of modern corals and sea water ($146.6\text{--}149.6\text{‰}$)²⁵. This left two strictly reliable ages from the lower reef and five strictly reliable ages from the upper reef (Fig. 2 and Supplementary Table 1). The two ages from the lower reef, 132 and 134 kyr, indicate that it is older than the upper reef, but, given the marginal ^{238}U concentrations, their accuracy is suspect. Ages from the upper reef show a range between 125 and 117 kyr, but the older ages are from clasts of *A. palmata* in the reef-flat zone, and could potentially be transported from the lower reef by hurricanes (thus making them unreliable for sea-level determination). The two strictly reliable subsamples from *in situ* *A. palmata* colonies in the reef crest (xD4-2a and xD4-3a) gave younger ages, between 117 and 119.5 kyr, and most likely indicate the true age of upper-reef tract development (Fig. 2). The only other *in situ* coral dated from the upper-reef crest (xE2-2a) also gave an age of 117.7 kyr, but this is accurate only to within 2 kyr.

Despite the lack of a good age from the lower-reef tract, differences in biofacies and elevation confirm that the two reefs are contemporaneous and had a back-stepping pattern of development. Progradation of the overlying beach unit during sea-level fall at the end of the highstand, for example, caused infiltration of an abraded shell gravel into the upper-reef crest and adjacent zones. This indicates that the reef had an open framework and was alive just before sea-level fell. No infiltration of beach gravel occurred in the lower-reef crest, however, because surface porosity was already occluded by a cap of crustose coralline algae (Fig. 1). This lack of infiltration requires that the lower reef be older, and the presence of a coralline cap signifies that

the lower reef be older, and the presence of a coralline cap signifies that

it died suddenly but remained submerged in sea water. Combined with its greater elevation, these differences imply that the upper reef back-stepped during sea-level rise.

The timing of reef demise and back-stepping is further constrained by key differences in the stratigraphic transition between the two reef tracts. In the lagoon, the transition is discontinuous and marked by a sharp, erosive contact across which there is a switch from patch-reef corals to sediment-tolerant intergrowths of branching coralline algae²⁶ and small-branched *Porites* (see D8 to D6, Fig. 1). By contrast, on the adjacent but more elevated windward edge of the patch-reef complex, the transition into the upper crest is abrupt but continuous, implying that conditions changed rapidly but remained conducive to continuous coral growth (see E4 to C4, Fig. 1). The sudden demise of reef-crest corals in the lower reef was therefore not only accompanied by sudden, partial demise in the lagoon, but also by a sudden ecological shift to reef-crest conditions along the windward edge of the patch-reef complex. This continuous development between reef tracts not only requires that they were contemporaneous but that the demise of the lower-reef crest at +3 m was ecologically synchronous with back-stepping and relocation of the upper-reef crest (Fig. 1c).

Furthermore, sea-level indicators demonstrate that this abrupt reef-crest back-stepping occurred at a sea-level position that was higher by 2–3 m. We assume that the maximum elevation of the lower-reef crest at +3 m closely represents mean low water and rule out the possibility that it was a submerged feature on the basis of the distinct breakwater facies zonation and the development of a lagoon where patch reefs reached the same elevation (Fig. 1c and Supplementary Fig. 1). The conclusion that this was a breakwater reef is also supported by the presence of a shallow, depth-restricted encruster association¹⁹ of coralline algae, *Homotrema rubrum* and vermetid gastropods on clasts and colonies of *A. palmata*.

The upper-reef crest, by contrast, contains clear indicators of a sea-level position higher than +3 m during its earliest development. *In situ* colonies of *A. palmata* up to 1.5 m tall are developed at the base of the reef-crest unit between +3 and +4.5 m (see D4 and E2, Fig. 1). If those colonies reached mean low water, then sea level must have been a minimum of +4.5 m when the crest developed. This higher sea-level position is also supported by the +4–4.5-m elevation of the base of the reef-flat deposit, which indicates that it could only have developed at a sea-level position greater than or equal to +4.5 m (see E4, D5 and C5, Supplementary Fig. 1). Finally, the reef-crest encruster association only appears above an elevation of +3.5 m,

which is consistent with a sea-level position of +5.5 m. Combined, these indicators provide reliable evidence for a minimum sea-level position of +5 m during early development of the upper-reef tract. We note, however, that the upper reef-flat surface at +5.5 m is close to the maximum reef-crest elevation of +5.8 m. The possibility that sea level jumped from +3 m to its final highstand elevation of +6 m cannot therefore be excluded (Fig. 1).

Given the absence of widespread reef development above +3 m in stable, low-lying areas like the Bahamas and Western Australia, the complete reef section at Xcaret provides a key insight into the final stage of reef development during the last interglacial. Reef demise and back-stepping at Xcaret was not only accompanied by localized erosion in the lagoon, but also by an abrupt switch to a sediment-tolerant assemblage²⁶. This implies that the 2–3-m sea-level jump outpaced accretion of the lower-reef crest and created a higher-energy wave field that enhanced sediment flux and smothered and eroded corals in lagoonal depressions. Coral growth could only continue along the elevated edge of the patch-reef complex and quickly developed large colonies of *A. palmata*, which formed the core of the upper-reef crest. This single jump in sea-level was therefore responsible for reef demise, marine erosion, back stepping and the suppression of reef growth—features that are all common along other low-lying coasts and that have been used to support claims of multiple sea-level excursions during the last interglacial⁷ (Supplementary Discussion).

To constrain the timing of the sea-level jump during the last interglacial, we screened isotopically reliable ages from the Bahamas⁴ and Western Australia^{5,6} for stratigraphic consistency²⁷ (Methods) and compared these data with coral ages and reef-crest sea-level indicators from Xcaret (Fig. 3). We identified three stratigraphically reliable age groups from the extensive Western Australia data set which indicate that widespread reef development at +3 m occurred during a 5-kyr stillstand of sea level between 126 and 121 kyr ago (Fig. 3). Most of the reliable ages of corals from the Bahamas fall into the same interval but do not form stratigraphically consistent groups. Given the limited occurrence of reef development up to +6 m in these stable areas, and the younger age of *in situ* reef-crest corals from the upper reef at Xcaret, we tentatively place the timing of the sea-level jump that caused reef demise and back-stepping at Xcaret immediately after the youngest reliable age group, 121 kyr ago. Although the precision of these ages precludes any direct measurement of the rise rate involved in the jump, it was most likely similar to rates that caused ecologically sudden demise and back-stepping of Caribbean reefs during the last

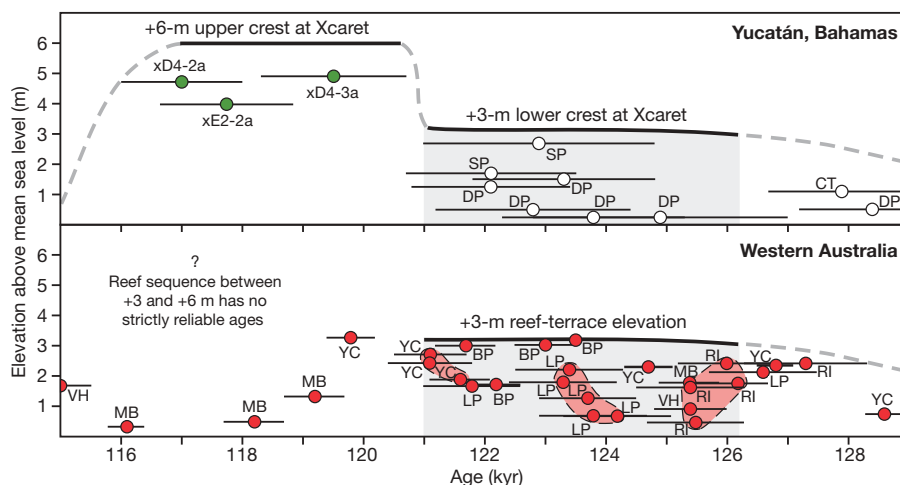


Figure 3 | Relative sea-level reconstructions for the last interglacial highstand. Open circles are isotopically reliable ages from the Bahamas⁴. (CT, Cockburn Town, San Salvador; SP, Sue Point, San Salvador; DP, Devil's Point, Great Inagua). Green circles are isotopically reliable ages from *in situ* upper-reef-crest corals at Xcaret. Sea-level position is defined by surface elevation of Xcaret reef-crest units. Red circles are isotopically reliable ages

from Western Australia^{5,6}. (VH, Vlaming Head; MB, Mangrove Bay; YC, Yardie Creek; BP, Burney Point; LP, Leander Point; RI, Rottneest Island.) Sea-level position is defined by surface elevation of the reef terrace. Stratigraphically consistent age groups highlighted in Western Australia data define a 5-kyr interval for the +3-m sea-level stillstand (shaded interval). Error bars, 2σ .

deglaciation^{10,12}. During those jumps, direct measurement of rise rates shows that they exceeded 36 mm yr^{-1} (refs 10, 28; see Supplementary Fig. 3). Our discovery of an ecologically sudden demise and back-stepping signature in reef-crest deposits from the Yucatán is therefore compelling evidence for a sea-level jump with a similar rise rate during the late stages of the last interglacial. This jump implies that an episode of ice-sheet instability, characterized by rapid ice loss, occurred late during an interglaciation that was warmer than present.

In our warming world, the implications of a rapid, metre-scale sea-level jump late during the last interglacial are clear for both future ice-sheet stability and reef development. Given the dramatic disintegration of ice shelves² and discovery of rapid ice loss from both the Antarctic and Greenland ice sheets³, the potential for sustained rapid ice loss and catastrophic sea-level rise in the near future is confirmed by our discovery of sea-level instability at the close of the last interglacial. Furthermore, the inhibition of reef development that this instability caused has negative implications for the future viability of modern reefs, which are already being impacted by anthropogenic activity on a global scale²⁹.

METHODS SUMMARY

We characterized detrital, framework and encruster facies in reef stratigraphic sections following ref. 19, and recorded contact types and elevations using a sea-level datum ($\pm 15 \text{ cm}$). Lateral continuity of units was physically traced between sections.

U-series measurements on coral samples used multistatic, multi-ion-counting inductively coupled plasma mass spectroscopy, following ref. 30. Whole-procedure blank values of the measured sample set were $\sim 2 \text{ pg}$ for Th and $4\text{--}8 \text{ pg}$ for U, both typical of this method and laboratory.

We based isotopic screening for potential U and Th loss or gain on several standard criteria⁸: the ^{238}U concentration should reflect modern coral species values ($2.0\text{--}3.5 \text{ p.p.m.}$), the ^{232}Th concentration should be $< 2 \text{ p.p.b.}$ and the abundance of calcite must be below X-ray diffraction detection limits. For samples meeting these criteria, age reliability is based on the $\delta^{234}\text{U}(T)$ criterion. We considered ages with the same $\delta^{234}\text{U}(T)$ value as modern corals and sea water ($146.6\text{--}149.6\text{‰}$)²⁵ to be strictly reliable to within the analytical uncertainty; ages with values of $149 \pm 8\text{‰}$ were considered to be accurate to $\pm 2 \text{ kyr}$ (refs 4, 6, 13, 14), and ages with values $> 149 \pm 8\text{‰}$ were considered to be unreliable.

Combined ^{230}Th and ^{231}Pa dating has shown that the $\delta^{234}\text{U}(T)$ criterion alone is insufficient to identify all corals affected by open-system diagenesis⁸. Quantification of extrinsic age variability in Holocene reefs has recently provided a criterion for screening isotopically reliable ages for stratigraphic consistency²⁷. This is based on the finding that a minimum of three corals growing within a 3-m vertical stratigraphic interval should not have an age distribution of $> 1 \text{ kyr}$. As a result, we consider the ages of groups of three or more corals that vary by no more than 1 kyr and come from conformable stratigraphic sections no thicker than 3 m to be stratigraphically reliable.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 October 2008; accepted 23 February 2009.

- Overpeck, J. T. *et al.* Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise. *Science* **311**, 1747–1750 (2006).
- Thomas, R. *et al.* Accelerated sea-level rise from West Antarctica. *Science* **306**, 255–258 (2004).
- Bamber, J. L., Alley, R. B. & Joughin, I. Rapid response of modern day ice sheets to external forcing. *Earth Planet. Sci. Lett.* **257**, 1–13 (2007).
- Chen, J. H., Curran, H. A., White, B. & Wasserburg, G. J. Precise chronology of the last interglacial period: ^{234}U – ^{230}Th data from fossil coral reefs in the Bahamas. *Geol. Soc. Am. Bull.* **103**, 82–97 (1991).
- Stirling, C., Esat, T., McCulloch, M. & Lambeck, K. High-precision U-series dating of corals from Western Australia and implications for the timing and duration of the last interglacial. *Earth Planet. Sci. Lett.* **135**, 115–130 (1995).
- Stirling, C., Esat, T., Lambeck, K. & McCulloch, M. Timing and duration of the last interglacial: evidence for a restricted interval of widespread coral reef growth. *Earth Planet. Sci. Lett.* **160**, 745–762 (1998).
- Neumann, A. C. & Hearty, P. J. Rapid sea-level changes at the close of the last interglacial (substage 5e) recorded in Bahamian island geology. *Geology* **24**, 775–778 (1996).
- Scholz, D. & Mangini, A. How precise are U-series coral ages? *Geochim. Cosmochim. Acta* **71**, 1935–1948 (2007).
- Fairbanks, R. G. A 17,000-year long glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Blanchon, P. & Shaw, J. Reef drowning during the last deglaciation: evidence for catastrophic sea-level rise and ice-sheet collapse. *Geology* **23**, 4–8 (1995).
- Bard, E. *et al.* Deglacial sea-level record from Tahiti corals and the timing of global meltwater discharge. *Nature* **382**, 241–244 (1996).
- Blanchon, P., Jones, B. & Ford, D. C. Discovery of a submerged relic reef and shoreline off Grand Cayman: further support for an early Holocene jump in sea level. *Sedim. Geol.* **147**, 253–270 (2002).
- Esat, T., McCulloch, M., Chappell, J., Pillans, B. & Omura, A. Rapid fluctuations in sea level recorded at Huon Peninsula during the penultimate deglaciation. *Science* **283**, 197–201 (1999).
- Stein, M. *et al.* TIMS U-series dating and stable isotopes of the last interglacial event in Papua New Guinea. *Geochim. Cosmochim. Acta* **57**, 2541–2554 (1993).
- Bruggemann, J. H. *et al.* Stratigraphy, palaeoenvironments and model for the deposition of the Abdur Reef Limestone: context for an important archaeological site from the last interglacial on the Red Sea coast of Eritrea. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **203**, 179–206 (2004).
- Thompson, W. G. & Goldstein, S. L. Open-system coral ages reveal persistent suborbital sea-level cycles. *Science* **308**, 401–404 (2005).
- Rohling, E. J. *et al.* High rates of sea-level rise during the last interglacial period. *Nature Geosci.* **1**, 38–42 (2008).
- Wilson, M., Curran, H. & White, B. Paleontological evidence of a brief global sea-level event during the last interglacial. *Lethaia* **31**, 241–250 (1998).
- Blanchon, P. & Perry, C. T. Taphonomic differentiation of *Acropora palmata* facies in cores from Campeche Bank Reefs, Gulf of Mexico. *Sedimentology* **51**, 53–76 (2004).
- Aronson, R. B., Macintyre, I. G. & Precht, W. F. Event preservation in lagoonal reef systems. *Geology* **33**, 717–720 (2005).
- Zúñiga, F. R., Reyes, M. A. & Valdés, C. A general overview of the catalog of recent seismicity compiled by the Mexican Seismological Survey. *Geofis. Int.* **39**, 161–170 (2000).
- Marquez-Azua, B., Cabral-Cano, E., Correa-Mora, F. & DeMets, C. A model for Mexican neotectonics based on nationwide GPS measurements, 1993–2001. *Geofis. Int.* **43**, 319–330 (2004).
- Szabo, B. J., Ward, W. C., Weidie, A. E. & Brady, M. J. Age and magnitude of the late Pleistocene sea-level rise on the eastern Yucatan Peninsula. *Geology* **6**, 398–406 (1978).
- Potter, E.-K. & Lambeck, K. Reconciliation of sea-level observations in the Western North Atlantic during the last glacial cycle. *Earth Planet. Sci. Lett.* **217**, 171–181 (2003).
- Delanghe, D., Bard, E. & Hamelin, B. New TIMS constraints on the uranium-238 and uranium-234 in seawaters from the main ocean basins and the Mediterranean Sea. *Mar. Chem.* **80**, 79–93 (2002).
- Steneck, R., Macintyre, I. & Reid, R. A unique algal ridge system in the Exuma Cays, Bahamas. *Coral Reefs* **16**, 29–37 (1997).
- Edinger, E. N., Burr, G. S., Pandolfi, J. M. & Ortiz, J. C. Age accuracy and resolution of Quaternary corals used as proxies for sea level. *Earth Planet. Sci. Lett.* **253**, 37–49 (2007).
- Peltier, W. R. & Fairbanks, R. G. Global glacial ice volume and Last Glacial Maximum duration from an extended Barbados sea level record. *Quat. Sci. Rev.* **25**, 3322–3337 (2006).
- Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742 (2007).
- Fietzke, J., Liebetrau, V., Eisenhauer, A. & Dullo, C. Determination of uranium isotope ratios by multi-static MIC-ICP-MS: method and implementation for precise U- and Th-series isotope measurements. *J. Anal. At. Spectrom.* **20**, 395–401 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Sanchez, E. Rios and R. Raigoza for providing park access and hospitality; E. Jordan-Dahlgren for original descriptions of the aquarium trench; G. Jordan-Garza and I. Pino for field assistance; A. Kolevica and K. Fischer von Mollard for help with the ^{230}Th measurements; and C. Neumann, P. Hearty, C. Holmden, D. Hopley, A. Curran and W. C. Ward for discussions. Funding was provided by DGAPA grant IN218799.

Author Contributions P.B. initiated the study, collected and analysed the field data and wrote the manuscript. A.E., J.F. and V.L. performed the U-series and X-ray diffraction analyses, and A.E. discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.B. (blanchon@icmyl.unam.mx).

METHODS

Stratigraphy. Reef stratigraphic sections were measured using a sea-level datum and thus have a potential elevation error of ± 15 cm corresponding to the spring tide amplitude. (That is, all sections intersected the water table, which has a hydraulic gradient of $7\text{--}10\text{ mm km}^{-1}$ (ref. 31). In coastal areas such as Xcaret, it is equivalent to mean sea level and oscillates with the tidal cycle with no apparent lag³¹.) Sections were logged by differentiating detrital carbonate facies from framework facies and recording their contact types and elevations. We performed characterization of detrital and framework facies following ref. 19. The lateral continuity of these units was physically traced between sections using an extensive network of Park tunnels and excavations. Full data on the sedimentology and stratigraphy will be published in a companion paper.

Chronology. U-series measurements of coral ages were performed at the Leibniz Institute of Marine Sciences at the University of Kiel (IFM-GEOMAR). Element separation procedure follows previously published methods but used Eichrom-UTEVA resin³⁰. Determination of U and Th isotope ratios followed a multistatic, multi-ion-counting inductively coupled plasma mass spectroscopy approach³⁰. For isotope dilution measurements, a combined $^{233}\text{U}/^{236}\text{U}/^{229}\text{Th}$ spike was used, with stock solutions calibrated for concentration using NIST-SRM 3164 (U) and NIST-SRM 3159 (Th) and, as combi-spike, calibrated against CRM-145 uranium standard solution (also known as NBL-112A) for U isotope composition and against a secular equilibrium standard (HU-1, uranium ore solution) for the precise determination of $^{230}\text{Th}/^{234}\text{U}$ activity ratios. Whole-procedure blank values of this sample set were measured to be around 2 pg for Th and between 4 and 8 pg for U. Both values are in the range typical of this method and the laboratory.

Isotopic screening for potential U and Th loss or gain was based on several standard criteria⁸: the calculated $\delta^{234}\text{U}(T)$ values should lie within the range of modern corals and sea-water, between $146.6 \pm 1.4\text{‰}$ and $149.6 \pm 1.0\text{‰}$ (ref. 25); the ^{238}U concentration should reflect modern coral species values of between 2.0 and 3.5 p.p.m.; the ^{232}Th concentration should be <2 p.p.b.; and the abundance of calcite must be below X-ray diffraction detection limits ($<1\%$ calcite). For

samples within the ^{238}U and ^{232}Th concentration ranges, and which lack detectable calcite, the reliability of coral ages is based on the $\delta^{234}\text{U}(T)$ criterion: ages with the same $\delta^{234}\text{U}(T)$ value as pristine modern coral and sea water ($146.6\text{--}149.6\text{‰}$) are considered to be strictly reliable to within the analytical uncertainty (assuming that the marine $^{234}\text{U}/^{238}\text{U}$ ratio has remained constant³²), whereas ages with values of $149 \pm 8\text{‰}$ are considered to be accurate to $\pm 2\text{ kyr}$ ^{4,6,13,14} and ages with values that exceed $149 \pm 8\text{‰}$ are considered to be unreliable.

Combined ^{230}Th and ^{231}Pa dating has shown that the $\delta^{234}\text{U}(T)$ criterion alone is insufficient to identify all corals affected by open-system diagenesis and that $>50\%$ of ^{230}Th ages with reliable isotopic values can have discordant ^{231}Pa ages⁸. An additional test with significant potential to screen coral ages for open-system behaviour is stratigraphic consistency. The expectation that ages in a conformable sequence should simply decrease with increasing elevation is unrealistic, as shallow reef-crest deposits are commonly mixtures of hurricane-emplaced fragments and *in situ* colonies¹⁹. Hurricanes therefore cause temporal mixing within individual reef-crest units as they develop, and those that have been dated at a sub-metre resolution show age reversals of up to 850 yrs^{12,19}. More rigorous quantification of such extrinsic age variability in Holocene reefs²⁷ has recently provided a criterion to better screen isotopically reliable ages for stratigraphic consistency. This criterion is based on the finding that a minimum of three corals growing within a 3-m vertical stratigraphic interval (that is, the average amount of accretion found to occur in 1 kyr or less) should not have an age distribution of >1 kyr (limit of age resolution possible with a minimum of three coral ages). As a result, we consider the ages of groups of three or more corals that vary by no more than 1 kyr and come from conformable stratigraphic sections no thicker than 3 m to be stratigraphically reliable.

31. Gonzalez-Herrera, R., Sanchez-y-Pinto, I. & Gamboa-Vargas, J. Groundwater-flow modeling in the Yucatan karstic aquifer, Mexico. *Hydrogeol. J.* **10**, 539–552 (2002).

32. Henderson, G. M. Seawater ($^{234}\text{U}/^{238}\text{U}$) during the last 800 thousand years. *Earth Planet. Sci. Lett.* **199**, 97–110 (2002).

Gene regulatory logic of dopamine neuron differentiation

Nuria Flames¹ & Oliver Hobert¹

Dopamine signalling regulates a variety of complex behaviours, and defects in dopamine neuron function or survival result in severe human pathologies, such as Parkinson's disease¹. The common denominator of all dopamine neurons is the expression of dopamine pathway genes, which code for a set of phylogenetically conserved proteins involved in dopamine synthesis and transport. Gene regulatory mechanisms that result in the direct activation of dopamine pathway genes and thereby ultimately determine the identity of dopamine neurons are poorly understood in all systems studied so far². Here we show that a simple *cis*-regulatory element, the dopamine (DA) motif, controls the expression of all dopamine pathway genes in all dopaminergic cell types in *Caenorhabditis elegans*. The DA motif is activated by the ETS transcription factor AST-1. Loss of *ast-1* results in the failure of all distinct dopaminergic neuronal subtypes to terminally differentiate. Ectopic expression of *ast-1* is sufficient to activate the dopamine pathway in some cellular contexts. Vertebrate dopamine pathway genes also contain phylogenetically conserved DA motifs that can be activated by the mouse ETS transcription factor Ets1 (also known as ER81), and a specific class of dopamine neurons fails to differentiate in mice lacking Ets1. Moreover, ectopic Ets1 expression induces dopaminergic fate marker expression in neuronal primary cultures. Mouse Ets1 can also functionally substitute for *ast-1* in *C. elegans*. Our studies reveal a simple and apparently conserved regulatory logic of dopamine neuron terminal differentiation and may provide new entry points into the diagnosis or therapy of conditions in which dopamine neurons are defective.

Nervous systems generally harbour distinct populations of dopamine (DA) neurons that derive from different precursor cells. Despite their diverse origin, all DA neurons share the expression of a core set of five genes that code for enzymes and transporters which synthesize, package and re-uptake dopamine ('dopamine pathway genes'; Fig. 1a). The regulatory logic of the terminal differentiation of DA neurons, manifested by the induction of the DA pathway genes, can, in theory, be described by two distinct models. In model 1, each dopamine pathway gene is independently activated by a distinct set of regulatory factors and, as a reflection of their distinct developmental history, each DA neuron subtype uses a distinct set of regulatory molecules (Fig. 1b). In model 2, each dopamine pathway gene is regulated by the same regulatory factor(s) and those factor(s) are the same in each DA neuron subtype (Fig. 1b). These two models make specific predictions about the *cis*-regulatory architecture of dopamine pathway genes. In model 1, each dopamine pathway gene is controlled by distinct *cis*-regulatory motifs and different motifs are active in individual DA neuron subtypes. In model 2, there is a single motif for all pathway genes that is used in all different DA neuron subtypes (Fig. 1b). To test these models, we made use of the DA system of the nematode *C. elegans*, which contains four distinct, lineally unrelated classes of DA neuron subtypes that express the same set of highly conserved

dopamine pathway genes (Fig. 1c)³. We systematically dissected the *cis*-regulatory regions of all DA pathway genes in the context of *gfp* reporters expressed in transgenic worms (Fig. 1c–h and Supplementary Fig. 1). The *cis*-regulatory analysis of two genes exclusively expressed in the DA neurons, the dopamine transporter gene *dat-1* (dopamine transporter) and the tyrosine hydroxylase gene *cat-2* (tyrosine hydroxylase), revealed the existence of a small *cis*-regulatory module (CRM) in each promoter that is required and sufficient to drive expression in all DA neurons (Fig. 1d, e). Dopamine pathway genes expressed in both DA and serotonergic (5-HT) neurons (*cat-1*, vesicular monoamine transporter, *cat-4*, GTP cyclohydrolase, *bas-1*, aromatic amino acid decarboxylase) contain separable CRMs for expression in DA and 5-HT neurons (Fig. 1f–h).

The DA-specific CRM of the *dat-1* locus contains a small sequence motif that is conserved in three other *Caenorhabditis* species (Fig. 1d and Supplementary Fig. 5); mutational analysis demonstrates that this motif is required for *dat-1* expression in all DA neurons (Fig. 1d). This motif was also sufficient to drive expression in all DA neurons, either when tested in isolation or when appended to the CRM of another neuron-specific gene (Supplementary Fig. 2a). Bioinformatics analysis predicted the binding of six different types of transcription factors to this conserved motif (Supplementary Fig. 2b). Point mutations that specifically abolished the predicted binding of some factors while keeping others intact revealed that the only predicted motif that can be made responsible for *cis*-regulatory motif activity in the DA neurons is a predicted ETS transcription factor binding site (EBS) defined by an invariant GGAW core sequence (Supplementary Fig. 2b). The DA-expressed CRMs of all other dopamine pathway genes also contain predicted EBSs, and mutational analysis corroborated their requirement for the correct expression in all DA neurons of *C. elegans* hermaphrodites (Fig. 1f–h and Supplementary Fig. 3) and in the three additional DA neuron pairs present in the male (Supplementary Fig. 4). All the functionally characterized EBSs are conserved in other *Caenorhabditis* species; they can occur in either orientation and at different distances from the transcriptional start (Supplementary Fig. 5). The weight matrix generated with all these sequences defines a consensus EBS sequence motif that we term the 'DA motif' (Fig. 1i and Supplementary Fig. 5).

By analysing the expression of the DA marker *dat-1::gfp* in mutants that lack each of the ten *C. elegans* ETS family members (Supplementary Fig. 6), we found that all *ets* family mutants showed wild-type *dat-1::gfp* expression except for animals lacking the *axon steering defect-1* (*ast-1*) gene, previously identified as a gene controlling axon outgrowth in the ventral nerve cord⁴. Moreover, we found that a mutant allele, *ot417*, which we retrieved from an unbiased forward genetic screen for mutants in which DA fate is inappropriately executed⁵, is an allele of *ast-1* (Fig. 2a). The expression of all five dopamine pathway genes was strongly affected if not completely lost in *ast-1* mutants (Fig. 2b, c, Supplementary Table 1 and Supplementary

¹Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, New York 10032, USA.

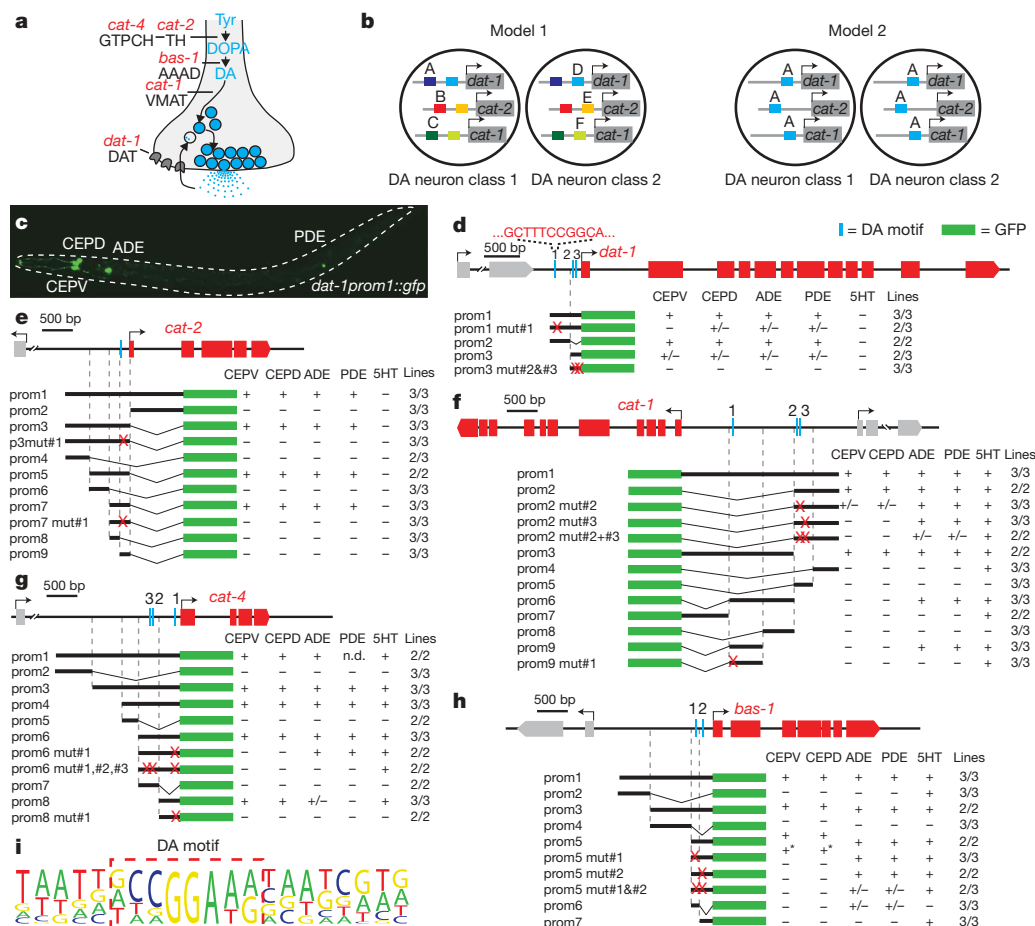


Figure 1 | Characterization of the DA motif in *C. elegans*. **a**, Schematic representation of a DA neuron synapse. AAAD, aromatic L-amino acid decarboxylase; DOPA, 3,4-dihydroxy-L-phenylalanine; Tyr, tyrosine. **b**, Schematic representation of two different models for DA terminal differentiation. See text for explanations. Capital letters represent hypothetical *trans*-acting factors. **c**, Picture of an adult worm expressing GFP under the control of the full-length *dat-1* promoter, labelling all *C. elegans* DA neurons. Similarly, *cat-2* is also exclusively expressed in DA neurons (data not shown), as *C. elegans* contains no adrenergic or noradrenergic neurons. **d**, *dat-1* promoter analysis. Schematic representation of the *dat-1* locus with its upstream region. Exons are represented as red blocks; the upstream gene is shown in grey. At the bottom is a representation of cloned and injected constructs, and the expression pattern in the DA and 5-HT neurons of transgenic worms. Thick black lines

symbolize the promoter piece placed in front of GFP (green box). Red cross represents a mutated EBS. '+' indicates >10% penetrant expression in more than half of the transgenic lines examined; '+/-' also means >10% penetrant expression, but the penetrance is lower than in the corresponding full-length construct; '-' indicates <10% penetrant expression in more than half of the transgenic lines examined; n.d., not determined. **e-h**, Analysis of the regulatory regions of all other dopamine pathway genes. +* means dimmer GFP expression than the corresponding wild-type construct. See Supplementary Figs 1 and 3 for all primary data and the nature of the mutations. **i**, The sequence alignment of all functional EBSs defines a position weight matrix (PWM) that is represented by a sequence logo. The conserved core in all sequences constitutes the DA motif. See Supplementary Fig. 5 for sequences used to define the DA motif.

Fig. 7). Two other DA terminal differentiation markers, the ion channels *asic-1* (ref. 6) and *trp-4* (ref. 7), also failed to be expressed in the DA neurons of *ast-1* mutants (Supplementary Fig. 8). Both genes contain phylogenetically conserved DA motifs in their regulatory regions. *ast-1* therefore appears to affect DA fate broadly, which is further corroborated by the axon pathfinding defects of DA neurons we observed in *ast-1* mutants (Supplementary Fig. 9). Loss of DA marker gene expression is not a reflection of early lineage specification defects and/or absence of the neurons, as assessed by analysis of additional fate markers (Supplementary Fig. 8).

ast-1 is expressed in several neurons⁴, including all DA neurons (Fig. 2d), and acts cell-autonomously in DA neurons, as the *ast-1* mutant phenotype can be rescued by expression of *ast-1* specifically in the DA neurons (Fig. 2e). *ast-1* expression persists in DA neurons throughout postembryonic stages, indicating that *ast-1* is required not only to initiate DA terminal cell fate but also to maintain DA neuron identity, a notion we confirmed through temporally controlled addition and removal of *ast-1* gene activity (Fig. 2f and Supplementary Fig. 10).

To address whether *ast-1* function is not only necessary for proper DA neuron differentiation but also sufficient, we ectopically induced *ast-1* expression throughout all cell and tissue types at different stages of development (Fig. 3). Ectopic induction during embryogenesis led to a substantial ectopic expression of both *dat-1::gfp* (Fig. 3a-d) and *cat-2::gfp* (data not shown). The morphology, location and pan-neuronal fate marker expression of these cells indicates that the effects of *ast-1* are confined to the nervous system, in which some (20 cells; ~10% of the embryonic nervous system) but clearly not all cells can be induced to ectopically express both *dat-1*/DAT and *cat-2*/TH. Ectopic *ast-1* was maximally effective when expressed around the time of neuronal differentiation (Supplementary Fig. 11). Moreover, ectopic *ast-1* expression under control of the DA- and 5-HT-specific *bas-1* promoter induced *dat-1*/DAT expression in 5-HT neurons (Fig. 3e) demonstrating that *ast-1* acts autonomously to control DA neuron specification and that 5-HT neurons provide the appropriate cellular context to allow *ast-1* to induce DA neuron specification. The related ETS domain transcription factor LIN-1 was not able to induce ectopic

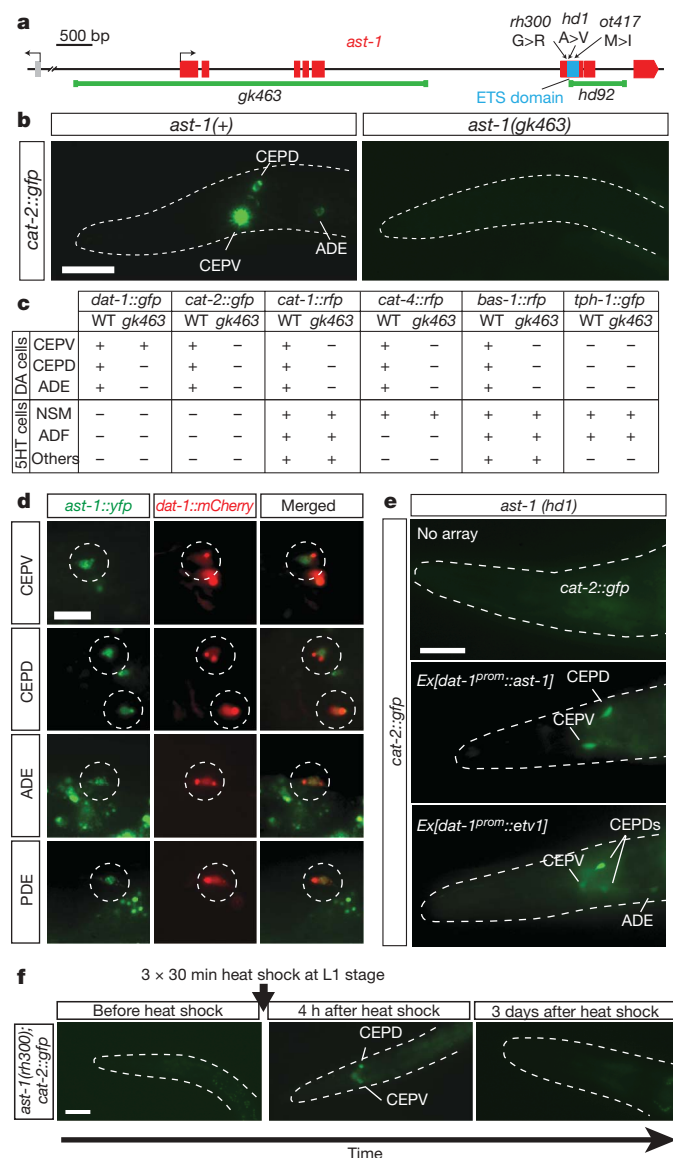


Figure 2 | *ast-1* is required to induce and maintain DA neuron differentiation. **a**, Schematic representation of the *ast-1* locus and mutant alleles available for this gene. **b**, Representative example of loss of DA fate marker in *ast-1* mutants. See Supplementary Figs 7 and 8 for other examples and Supplementary Table 1 for quantification of data. Scale bar, 10 μ m. **c**, Summary of *ast-1* null mutant phenotype. +, fate marker expressed; -, fate marker not expressed. Owing to early larval lethality, only the embryonically generated DA head neurons, but not the postembryonically generated PDE neurons, could be scored for developmental defects in *ast-1* null mutants. Markers that were expressed in both DA and 5-HT neurons were assayed with an *rfp* reporter in a transgenic background in which 5-HT neurons were labelled with *gfp* (*Is[tph-1::gfp]*) to allow scoring of loss of expression specifically in the DA neurons. See Supplementary Fig. 7 for primary data. WT, wild type. **d**, Expression of an *ast-1::yfp* reporter gene⁴ in DA neurons. DA neurons are labelled with *dat-1::mCherry*. Scale bar, 5 μ m. **e**, Rescue of the *ast-1* mutant phenotype. We used a hypomorphic allele, *hd1*, in which *dat-1* expression is unaffected (Supplementary Table 1), to drive *ast-1* or a mouse homologue, *Etv1*, under control of the DA-specific *dat-1* promoter and assayed expression of *cat-2::gfp* (*otIs199*). Scale bar, 8 μ m. **f**, Developmentally staged *ast-1(rh300)* animals, containing the heat-shock-inducible *ast-1* array *otIs198* and the DA fate marker *cat-2::gfp* (*otIs199*), were grown under non-inducible conditions to the first larval stage (resulting in an absence of *cat-2::gfp* expression in 100% of animals); *ast-1* was then induced by heat shock at the L1 stage. Of the 40 animals found to turn on expression of *cat-2::gfp* 4 h after heat shock, 17 lost expression after 3 days and all lost expression after 5 days. Data with a temperature-sensitive allele of *ast-1* corroborate sustained *ast-1* activity (Supplementary Fig. 10). Scale bar, 10 μ m.

DA neuron production when expressed under similar condition (data not shown), demonstrating the specificity of AST-1 function.

To assess whether ETS transcription factor(s) have a similar function in vertebrate DA neuron specification, we analysed their expression in the DA areas of the brain (Supplementary Table 2). Distinct ETS factors appeared to be expressed in distinct types of DA neurons and we focus here on the ETS factor *Etv1*, which is expressed in the DA neurons of the olfactory bulb (Supplementary Fig. 12)^{8,9}. Mice lacking *Etv1* (ref. 10) displayed a notable reduction in the number of tyrosine hydroxylase (TH)-positive cells in their olfactory bulb compared to wild-type siblings, whereas other periglomerular interneuron subtypes were not affected or were less severely reduced (Fig. 4a–c and Supplementary Fig. 13). This phenotype was not paralleled by increased cell death, by a reduction in the overall density of cells in the glomerular layer, by a reduction in overall neuron number or by proliferation defects (Fig. 4d–f and Supplementary Fig. 14). Moreover, the identity of DA progenitor cells in the lateral ganglionic eminence, which already express *Etv1* (ref. 11), appeared unaffected in *Etv1* mutants (Supplementary Fig. 15). Therefore, *Etv1* may affect a late stage in olfactory DA neuron differentiation.

Like *ast-1*, *Etv1* appears not only required for DA neuron differentiation but also sufficient, because ectopic expression of *Etv1* in olfactory bulb primary cell culture increases the number of cells expressing the DA marker TH (Fig. 4g). *Etv1* is also able to activate directly the *cis*-regulatory region of the mouse TH locus in a heterologous context (Fig. 4h). This activation depends on the presence of two phylogenetically conserved DA motifs (Fig. 4h). Like in *C. elegans*, phylogenetically conserved DA motifs can also be found in the 5' upstream regulatory region of all four other mouse dopamine pathway genes (Supplementary Fig. 16). Another indicator for a conserved function of mouse *Etv1* and worm AST-1 is that mouse *Etv1* is able to rescue the *ast-1* mutant phenotype when expressed in transgenic worms (Fig. 2e).

In conclusion, we have described here a surprisingly simple regulatory logic for DA specification. Our *cis*-regulatory analysis in worms reveals that all dopamine pathway genes are co-regulated through a similar *cis*-regulatory motif and *trans*-acting factor, and this regulatory logic applies to DA neurons of distinct lineage origin. Our analysis demonstrates that the ETS factor *ast-1* is a terminal selector gene for DA cell fate, akin to other terminal selector genes that control the terminal identity of other neuron types^{12–15}. Terminal selector genes are transcription factors that directly regulate the 'nut-and-bolts' differentiation gene batteries that determine the specific properties of a neuron by binding to simple *cis*-regulatory motifs shared by members of the terminal gene batteries, termed 'terminal selector motifs' (or, in the case of the DA neurons, the DA motif)¹⁵. As exemplified by AST-1, terminal selector genes are continuously expressed throughout the life of a neuron to ensure that the terminal differentiation state is properly maintained.

The regulatory logic of DA neuron specification seems to be phylogenetically conserved. Vertebrate dopamine pathway genes also contain DA motifs that are required for the activation by a *trans*-acting factor that is homologous to the *C. elegans* *trans*-acting factor. Loss of the *trans*-acting factor either in worms or in mice leads to a loss of the DA phenotype. Both *Etv1* and *ast-1* are continuously expressed throughout the postmitotic life of DA neurons, and our analysis in worms indicates that these factors also maintain the terminal identity of DA neurons. The function of vertebrate ETS proteins in DA specification may have been distributed over several different ETS domain transcription factors, as *Etv1* is not expressed in other DA neuron populations in the brain and as it does not affect the generation of these other types of DA neurons (data not shown). Those other areas express a related ETS factor, *Etv5*, which may fulfil a role similar to that of *Etv1* in olfactory DA neurons; in support of this notion, *Etv5* can also transactivate the TH promoter in a heterologous assay system (Supplementary Fig. 17). The logic of distributing an ancestral gene function, observed in an invertebrate species, over several vertebrate

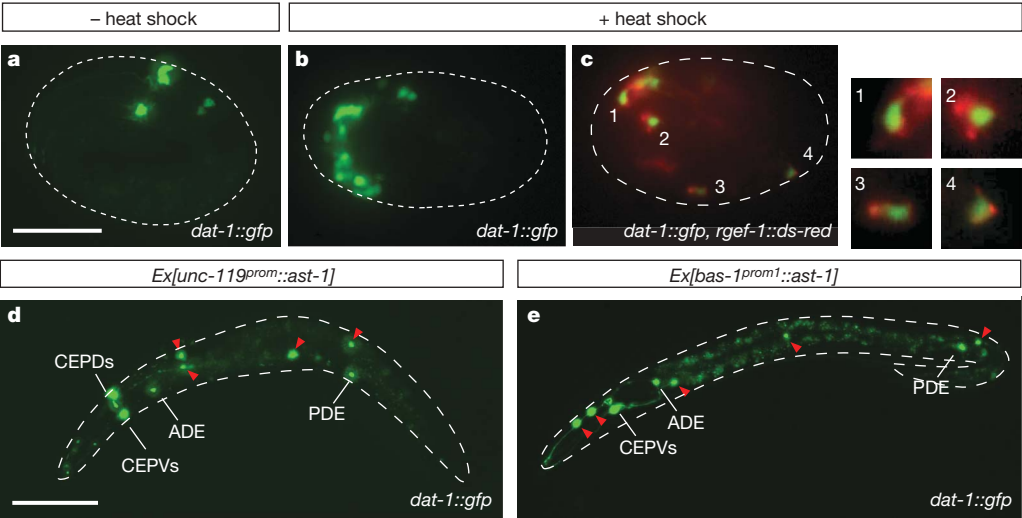


Figure 3 | Ectopic *ast-1* expression can induce DA cell fate. **a**, Representative control embryo after the threefold stage. *dat-1::gfp* expression starts at late threefold stage and can be detected in the six embryonically generated DA neurons. **b**, **c**, Representative picture of an embryo heat shocked 4 h after the two-cell stage and analysed 10 h after the heat shock. *dat-1::gfp* is ectopically expressed in many cells of the embryo. Scale bar, 20 μ m. In the presence of a pan-neuronal marker (*rgef-1::rfp*; **c**), the cells ectopically expressing DA fate can be identified as neurons. **d**, **e**, Ectopic expression of *ast-1* under the control of the ectodermal promoter *unc-119* and the DA/5HT-neuron-specific promoter *bas-1* leads to ectopic expression of *dat-1::gfp* in additional neurons compared to wild-type worms (red arrowheads). Similar effects were observed in multiple lines (2/2 lines for *bas-1* driver; 2/3 lines for *unc-119* driver). Scale bar, 100 μ m.

paralogues of the ancestral invertebrate orthologue has been noted for other transcription factors as well¹⁶ and seems an important component of driving neuronal diversification processes in more complex brains.

AST-1 and Etv1 both act as terminal selector genes for DA terminal differentiation, but their presence is not strictly sufficient to activate DA genes because both AST-1 and Etv1 are also expressed in neurons

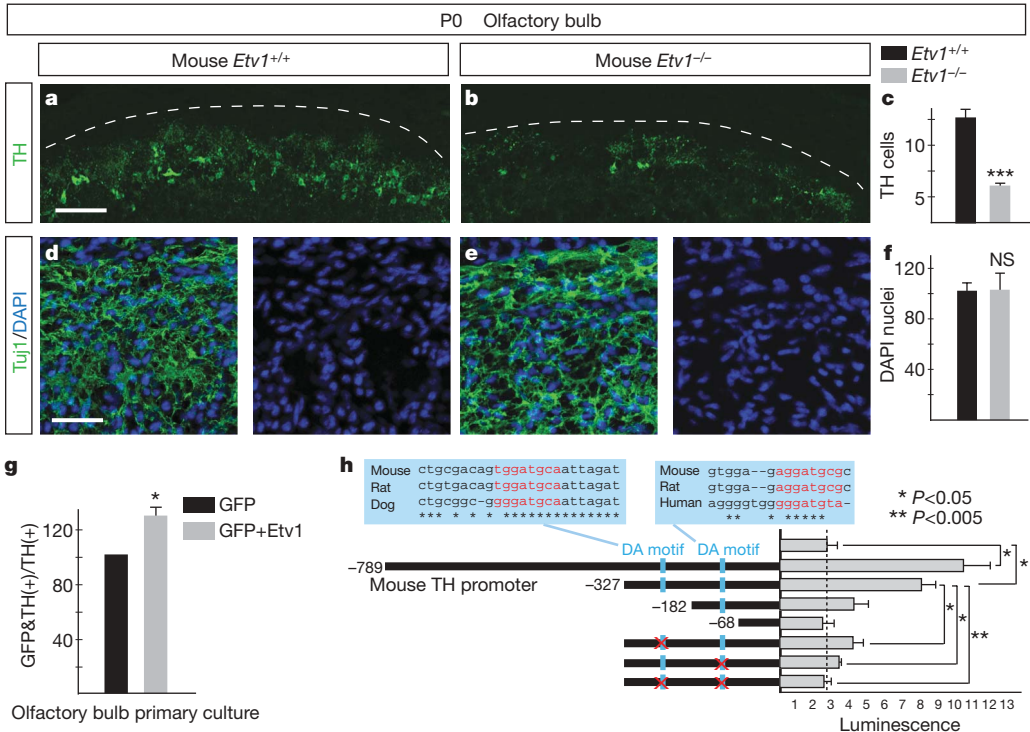


Figure 4 | Mouse *Etv1* is necessary for the olfactory bulb DA neuron specification **a**, **b**, Coronal section TH immunostaining of a wild-type (a) and *Etv1* mutant (b) newborn pup (P0) olfactory bulb. Scale bar, 150 μ m. **c**, Quantification of TH-positive cells in wild-type and *Etv1* mutants at P0. *Etv1* mutants show a significant reduction in the number of TH-positive cells already at this stage ($n = 3$, *** $P = 0.00009$). **d**, **e**, Coronal section TuJ1 immunostaining (green) and DAPI staining (blue) of a wild-type (d) and *Etv1* mutant (e) P0 glomerular layer to label neurons and cell nuclei, respectively. Scale bar, 40 μ m. **f**, Quantification of DAPI nuclei in wild-type and *Etv1* mutants at P0. Glomerular layer cell density is similar between wild-type and *Etv1* mutants ($n = 3$, $P = 0.93$). NS, not significant. **g**, Overexpression of Etv1 can induce DA differentiation. Dissociated P0 olfactory bulbs were transfected with *gfp* and vector control or *gfp* and *Etv1* expression vector, plated and cultured for 4 days. * $P < 0.05$. **h**, Analysis of the activation of TH promoter by Etv1 in COS cells. The dotted line indicates the level of luciferase activation of the empty luciferase vector observed on Etv1 transfection. DA motif denotes the phylogenetically conserved match to the VGGAWRNV consensus. In the top panel, asterisks denote conserved nucleotides. $n = 3$ independent experiments for each construct. All error bars represent s.e.m.

other than DA neurons^{4,17}. Our ectopic expression experiments also show that ability of AST-1 to induce ectopic DA fate is restricted to some cellular and temporal contexts. Classic 'master regulators', such as fly *Eyeless* or mouse *MyoD* (also known as *MyoD1*), also show similar context-dependencies in their mode of action^{18,19}. AST-1 and *Etv1* function may be actively inhibited in cells 'refractory' to AST-1/*Etv1* activity. Alternatively, AST-1/*Etv1* function may require additional, cell-type-specific factors for appropriate function in DA neurons. Such 'combinatorial coding' mechanisms are a common theme in neuron type specification²⁰, and our identification of a conserved role of ETS factors as a central component of such a code is the first important step in decoding the regulatory logic of DA neuron specification. It will be interesting to see whether the additional specificity determinants of ETS factors are also conserved from worms to vertebrates.

METHODS SUMMARY

Transgenic and mutant *C. elegans* strains. Reporter gene constructs were generated by subcloning into the pPD95.75 backbone vector, which contains the *gfp* coding sequence and the *unc-54* 3' untranslated region. Mutagenesis and deletions were performed using the Quickchange II XL site-directed mutagenesis kit (Stratagene). Reporter constructs were injected into *otIs181(Is[dat-1::mCherry;ttx-3::mCherry])* to facilitate the identification of the DA cells. DNA was injected at 50 ng μl^{-1} using *rol-6* as an injection marker. For every construct, 30 or more animals were scored from at least 2 different transgenic lines. Ets mutant strains were obtained from the knockout consortia in Oklahoma, Vancouver and Tokyo. For the heat-shock experiments, a transgenic strain of the following genotype was generated: OH7546 [*otIs198(hsp16-2::ast-1;hsp16-2::NLS-mCherry;ttx-3::ds-red)*, *vtIs1(dat-1::gfp;rol6)*]. Heat-shock induction conditions, as well as a list of transgenic and mutants strains, are provided in the Methods.

Analysis of vertebrate *Etv1*. Standard histological protocols were used to analyse wild-type and *Etv1* mutant mouse samples. For ectopic *Etv1* expression, olfactory bulbs were dissected from wild-type mice at postnatal day zero to one, dissociated and electroporated using the Amaxa Nucleofector System and the mouse neuron kit, following the manufacturer's protocol. For the TH promoter analysis, COS cells were transfected with lipofectamine (Invitrogen) and luciferase activity was measured using the luciferase assay kit (Stratagene) and the β -galactosidase enzyme assay system (Promega). See Methods for full description of histology, cell culture and quantification methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 February; accepted 26 February 2009.

Published online 15 March 2009.

- Iversen, S. D. & Iversen, L. L. Dopamine: 50 years in perspective. *Trends Neurosci.* **30**, 188–193 (2007).
- Abeliovich, A. & Hammond, R. Midbrain dopamine neuron differentiation: factors and fates. *Dev. Biol.* **304**, 447–454 (2007).
- Nass, R. & Blakely, R. D. The *Caenorhabditis elegans* dopaminergic system: opportunities for insights into dopamine transport and neurodegeneration. *Annu. Rev. Pharmacol. Toxicol.* **43**, 521–544 (2003).
- Schmid, C., Schwarz, V. & Hutter, H. AST-1, a novel ETS-box transcription factor, controls axon guidance and pharynx development in *C. elegans*. *Dev. Biol.* **293**, 403–413 (2006).

- Doitsidou, M., Flames, N., Lee, A. C., Boyanov, A. & Hobert, O. Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nature Methods* **5**, 869–872 (2008).
- Vogl, G. & Tavernarakis, N. A synaptic DEG/ENAC ion channel mediates learning in *C. elegans* by facilitating dopamine signalling. *EMBO J.* **27**, 3288–3299 (2008).
- Li, W., Feng, Z., Sternberg, P. W. & Xu, X. Z. A. *C. elegans* stretch receptor neuron revealed by a mechanosensitive TRP channel homologue. *Nature* **440**, 684–687 (2006).
- Allen, Z. J. II, Waclaw, R. R., Colbert, M. C. & Campbell, K. Molecular identity of olfactory bulb interneurons: transcriptional codes of periglomerular neuron subtypes. *J. Mol. Histol.* **38**, 517–525 (2007).
- Saino-Saito, S. et al. ER81 and CaMKIV identify anatomically and phenotypically defined subsets of mouse olfactory bulb interneurons. *J. Comp. Neurol.* **502**, 485–496 (2007).
- Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene *Er81* controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).
- Flames, N. et al. Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
- Wenick, A. S. & Hobert, O. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev. Cell* **6**, 757–770 (2004).
- Duggan, A., Ma, C. & Chalfie, M. Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. *Development* **125**, 4107–4119 (1998).
- Etchberger, J. F. et al. The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev.* **21**, 1653–1674 (2007).
- Hobert, O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl Acad. Sci. USA* **105**, 20067–20071 (2008).
- Lichtneckert, R. & Reichert, H. Insights into the urbilaterian brain: conserved genetic patterning mechanisms in insect and vertebrate brain development. *Heredity* **94**, 465–477 (2005).
- Brown, T. A. & McKnight, S. L. Specificities of protein–protein and protein–DNA interaction of GABP alpha and two newly defined ets-related proteins. *Genes Dev.* **6**, 2502–2512 (1992).
- Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. *Science* **267**, 1788–1792 (1995).
- Weintraub, H. et al. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of *MyoD*. *Proc. Natl Acad. Sci. USA* **86**, 5434–5438 (1989).
- Shirasaki, R. & Pfaff, S. L. Transcriptional codes and the control of neuronal identity. *Annu. Rev. Neurosci.* **25**, 251–281 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Q. Chen for injection assistance, M. Doitsidou and J. Chen for providing the *ot417* allele, E. Savner for initiating some of the promoter analysis, A. Krüyer for technical help, the CGC for strains, the *C. elegans* Gene Knockout Consortia, led by S. Mitani at Tokyo Women's Medical University School of Medicine and by D. Moerman at the University of British Columbia, for knockout alleles, H. Hutter, N. Tavernarakis and T. Jessell for strains and reagents, several members of Columbia University community for sharing their equipment, Hobert laboratory members, especially M. Doitsidou, for discussion, and A. Abeliovich, R. J. Johnston and I. Greenwald for comments on the manuscript. This work was funded by the National Institutes of Health (R01NS039996-05; R01NS050266-03), the Howard Hughes Medical Institute and a EMBO long term fellowship and Marie Curie Outgoing International fellowship to N.F.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.F. (nf2171@columbia.edu) or O.H. (or38@columbia.edu).

METHODS

Strains and transgenes. For the ETS family analysis the following mutants were crossed into *vtIs1(dat-1::gfp)* strain to check for a DA phenotype: *tag-97(ok284)*; *C50A2.4(tm440)*; *C52B9.2(tm413)*; *ets-4(ok165)*; *C24A1.2(tm801)*; *ets-5(tm866, tm1734)*; *lin-1(e1777)* *ast-1(hd1, rh300, hd92, gk463)*. F19F10.1(*tm456*); F19F10.5(*tm436*) were crossed into *otIs181*. For further *ast-1* analysis the following strains were crossed into *ast-1(gk463 or hd92): nIs118(cat-2::gfp)*, *zIs13(tph-1::gfp)*, *oyIs59(osm-6::gfp)*, *AX43[lin15(n765);Ex(gcy-36::gfp)]*, *TQ886[trp-4::YFP+pha-1(+);pha-1(ts);him-5]*. The *cat-1^{prom1}::mCherry*, *bas-1^{prom1}::mCherry*, *cat-4^{prom1}::mCherry* and *asic-1::gfp* reporters were injected into the *ast-1* balanced strain *ast-1(gk463)* *bli-2(e768)* *unc-4(e120)/mIn1[mIs14 dpy-10(e128)]* II at 50 ng μl^{-1} . *AST-1* expression was analysed injecting *ast-1::yfp⁴* at 50 ng μl^{-1} into *otIs181*.

In regard to our reporter gene analysis, we note that we have previously reported that, in some cases, PCR products are expressed in more cells than the equivalent cloned product²¹. Some of the reporter gene constructs that gave no expression in the DA cells were injected both as subcloned plasmids and as PCR products (using vector primers). No difference in expression was found in any of them except for the single-copy 31 bp EBS motif from *dat-1* (Supplementary Fig. 2) that gave no expression injected as a clone product and PDE expression when injected as PCR fragment.

Heat-shock experiments. Strain OH7546 [*otIs198(hsp16-2::ast-1;hsp16-2::NLS-mCherry;ttx-3::ds-red)*, *vtIs1(dat-1::gfp;rol6)*] was used for the heat-shock experiments. For the rescue experiment shown in Fig. 2f (postembryonic heat shock), first larval stage worms were heat shocked at 37 °C three times for 30 min with 1 h incubation at 20 °C between each heat shock to let worms recover. After heat shocks, worms were kept at 15 °C and scored at the indicated times.

For the ectopic DA fate induction experiments (Fig. 3), two cell embryos were released from hermaphrodite mothers by sectioning them in half, mounted in slides, incubated at 20 °C, heat shocked at 37 °C for 20 min at different stages of development and analysed the following morning (approximately 24 h after first cleavage).

hsp16-2 promoter activity was monitored by *mCherry* expression from *hsp16-2::NLS-mCherry* construct present on the same array. Even though red fluorescence was first detected 2 h after the heat shock, immunodetection of protein production from the same promoter has been reported as early as 30 min after heat shock²².

Histology. P0 mice were anesthetized in ice and transcardially perfused with 4% paraformaldehyde (PFA). P10 mice were anesthetized with an overdose of ketamine HCl (Ketaset) and xylazine HCl (AmTech) and transcardially perfused with 4% PFA. Brains were removed, postfixed for 2 h at 4 °C, cryoprotected in 30% sucrose in PBS, embedded in Tissue-Tek OCT compound and stored frozen at -80 °C. 20- μm coronal sections were cut on a cryostat.

The following antibodies were used: sheep anti-tyrosine hydroxylase (1/1,000, Peel-Freeze), rabbit anti-Er81 (1/32,000, provided by T. Jessell), rabbit anti-calbindin (1/5,000, Swant), rabbit anti-calretinin (1/2,000, Chemicon), rabbit anti-caspase 3 cleaved (1/200, Cell Signaling), mouse anti- β -tubulin (Tuj1) (1/500 Covance), rabbit anti-phospho-histone H3 (1/500 Upstate).

Etv1 mutant analysis was performed using a confocal microscope (Leica TCS-SP5). For P0 tyrosine hydroxylase, P10 tyrosine hydroxylase, P10 calretinin and P10 calbindin olfactory bulb quantification, eight different fields corresponding to two different levels were scored in each animal. For P0 cleaved caspase 3 and phosphohistone 3 quantification, two whole sections of olfactory bulb were scored for each animal. For cell density quantification 4,6-diamidino-2-phenylindole (DAPI) nuclei were scored in a field of the glomerular layer for each animal. Olfactory bulb volume was estimated by the Cavalieri method²³. In brief, the whole extent of the olfactory bulb was cut in 20- μm coronal sections; one every twenty

sections was analysed using a grid of 0.04 mm². Data were statistically analysed using ANOVA tests.

In situ hybridization was performed as described before¹¹. CD-1 mouse embryos were analysed using the following probes: *Dlx2* (probe provided by J. Rubenstein); *Ehf* (image clone 5008262 and 3991773); *Elf3* (MTF387 from the Gray *et al.* library²⁴); *Elf5* (image clone 5252409); *Elk1* (image clone 423774); *Elk3* (image clone 3471706); *Elk4* (image clone 3589378); *Etv1* (probe provided by T. Jessell); *Erg* (image clone 7105647 and 7442975); *Ets1* (image clone 5720204); *Ets2* (image clone 3385111 and 3511332); *Etv3* (image clone 5290278); *Etv4* (antibody provided by T. Jessell); *Etv5* (image clone 3673190 and 3674281); *Etv6* (probe provided by S. H. Orkin); *Fli1* (image clone 717781 and 808955); *ngn2* (also known as *Neurog2*; probe provided by F. Guillemot); *Sfp1* (image clone 6529808 and 5003573); *Spib* (probe provided by M. C. Simon); *Spic*-C (also known as *Spig*; MTF416 from the Gray *et al.* library²⁴).

Olfactory bulb primary cultures. Olfactory bulb and cortex were dissected separately from 10–12 pups to obtain enough number of cells for the electroporation. Cells were dissociated and equal amounts of olfactory bulb and cortex cells were mixed to obtain approximately 6×10^6 cells per ml. Cells were electroporated using the Amaxa Nucleofector system and the mouse neuron kit, following the manufacturer's protocol. In brief, for each reaction 3×10^6 cells were electroporated with 0.7 μg of pmaxGFP, 2 μg of PCDNA3.1, 0.7 μg of pmaxGFP or 2 μg of *Etv1*/PCDNA3.1, respectively. Cells were immediately plated in 8-well glass chamber slides and cultured for 4 days as described previously²⁵. Cells were then fixed with 4% PFA and immunostained against with GFP and TH antisera. Two independent sets of experiments were performed and in each: two wells were scored for the GFP-expressing cells and two for the GFP+*Etv1*-expressing cells. In un-transfected cultures, the fraction of TH-positive cells was 1/3,000 plated cells. For each well we scored the total number of TH-positive cells and the number of co-labelled TH/GFP-positive cells (ranking from 100 to 200 TH cells and 20 to 50 TH/GFP cells per well). The percentage of TH transfected cells versus TH not transfected cells was calculated.

Luciferase assays. COS cells were plated in M6 plates and incubated overnight (12 h); at approximately 70% confluence cells were transfected with lipofectamine (Invitrogen). Each well was transfected with 2.7 μg of total DNA (200 ng of β -galactosidase, 500 ng of PGL3 luciferase, 2 μg of *Etv1*/PCDNA3.1(–) or PCDNA3.1(–)). Twenty-four hours after transfection, luciferase and β -galactosidase activity was measured using luciferase assay kit (Stratagene), β -galactosidase enzyme assay system (Promega) and 20/20ⁿ Luminometer (Turner BioSystems) and Biorad Model 680 Microplate reader. For each experiment each value represents the mean luciferase activity in three different wells and each construct was analysed in three independent experiments. Statistical analysis was performed applying the *t*-test.

DA motif mutagenesis was performed using the Quickchange II XL site-directed mutagenesis kit (Stratagene). In both cases the core GGAT of the DA motif was mutated to AGAT.

- Etchberger, J. F. & Hobert, O. Vector-free DNA constructs improve transgene expression in *C. elegans*. *Nature Methods* **5**, 3 (2008).
- Fukushige, T. & Krause, M. The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early *C. elegans* embryos. *Development* **132**, 1795–1805 (2005).
- Rosen, G. D. & Harry, J. D. Brain volume estimation from serial section measurements: a comparison of methodologies. *J. Neurosci. Methods* **35**, 115–124 (1990).
- Gray, P. A. *et al.* Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255–2257 (2004).
- Cobos, I., Borello, U. & Rubenstein, J. L. Dlx transcription factors promote migration through repression of axon and dendrite growth. *Neuron* **54**, 873–888 (2007).

LETTERS

Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates

Muriel C. Schneider^{1*†}, Beverly E. Prosser^{2*}, Joseph J. E. Caesar², Elisabeth Kugelberg¹, Su Li¹, Qian Zhang¹, Sadik Quoraishi², Janet E. Lovett², Janet E. Deane², Robert B. Sim³, Pietro Roversi², Steven Johnson², Christoph M. Tang¹ & Susan M. Lea²

The complement system is an essential component of the innate and acquired immune system¹, and consists of a series of proteolytic cascades that are initiated by the presence of microorganisms. In health, activation of complement is precisely controlled through membrane-bound and soluble plasma-regulatory proteins including complement factor H (fH; ref. 2), a 155 kDa protein composed of 20 domains (termed complement control protein repeats). Many pathogens have evolved the ability to avoid immune-killing by recruiting host complement regulators³ and several pathogens have adapted to avoid complement-mediated killing by sequestering fH to their surface⁴. Here we present the structure of a complement regulator in complex with its pathogen surface-protein ligand. This reveals how the important human pathogen *Neisseria meningitidis* subverts immune responses by mimicking the host, using protein instead of charged-carbohydrate chemistry to recruit the host complement regulator, fH. The structure also indicates the molecular basis of the host-specificity of the interaction between fH and the meningococcus, and informs attempts to develop novel therapeutics and vaccines.

Neisseria meningitidis is a human adapted pathogen of global importance as a leading cause of bacterial meningitis and septic shock⁵. Owing to Neisserial strain variation, the vaccines currently available for meningococcal disease are only effective against subsets of strains and do not provide universal protection^{6,7}. However, promising novel antigens have been identified⁸, including factor H binding protein (known as fHbp, GNA1870, or R2086), a 27 kDa surface lipoprotein which is present on the surface of all strains of *N. meningitidis* and elicits protective bactericidal antibodies^{9,10}. The structure of a portion of fHbp has been determined by NMR¹¹, and its function studied by subdividing it into a series of regions, termed 'A', 'B' and 'C'¹². fHbp is the sole receptor for fH on the meningococcus, and recruitment of fH contributes to the ability of the meningococcus to avoid innate immune responses by inhibiting complement-mediated lysis in human plasma^{13,14}. Individuals with polymorphisms in the promoter of the gene encoding fH (also known as CFH) that are associated with increased plasma fH levels are at increased risk from meningococcal disease¹⁵.

We have used a range of approaches to dissect the fH binding sites on fHbp (Fig. 1). High affinity interactions between fH and fHbp were only observed with fHbp constructs containing all three of the previously defined (see above) regions of fHbp (Fig. 1a), implying that fHbp has an extended recognition site for fH across its entire surface. We then sought to identify which of the 20 complement control protein (CCP) domains of fH mediate the interaction with

fHbp. Far-western, FACS (fluorescence-activated cell sorting; Fig. 1b) and SPR (surface plasmon resonance; Fig. 1c) analyses demonstrated that the key regions of fH recognized by fHbp are the sixth and seventh domains, CCPs 6 and 7 (called hereafter fH67) and that a construct containing both these domains was capable of inhibiting the fHbp-dependent interaction between fH and *N. meningitidis* (Fig. 1d). Quantification of the interaction demonstrates that the dissociation constant is $K_d \approx 5$ nM (Fig. 1c and e; Supplementary Table 1, Supplementary Fig. 1) for any construct containing CCPs 6 and 7. This interaction was not dissociated by high salt (1 M NaCl, not shown) or pH ranging from 4 to 8 (not shown), providing additional support for the high affinity nature of the binding event.

To better understand fH/fHbp recognition, we crystallized and determined crystal structures of the complex between fH67 and a fHbp construct comprising regions A, B and C but lacking the lipid anchor. Models for fHbp and fH67 were built and refined to 2.35 Å in two crystal forms with a total of seven independent copies of the complex (Fig. 2). The structure reveals that fHbp folds to form two β -barrels (Fig. 2a) with the amino-terminal barrel consisting of the A and part of the B regions, while the carboxy-terminal barrel (as previously seen in the NMR structure of the BC region¹¹, overlaid in Supplementary Fig. 2) is composed of the rest of the B and the C regions. Searching structural databases with the N-terminal barrel reveals no close structural homologues (Supplementary Analyses), and the distinct topologies of the β -barrels (Fig. 2c) suggest that they have not arisen by a gene duplication event.

The fH67–fHbp complex is held together by extensive interactions between both β -barrels of fHbp and fH CCP6 (Fig. 2b), with more minor contacts to CCP7, consistent with our binding studies. In particular, the helical insertion into the second β -strand of CCP6 (an unusual feature of this CCP domain) is centrally located in the complex. Analysis by the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) gives all seven independent complexes a significance score of 1.0 (that is, extremely likely to be biologically significant), with the average surface area of fHbp buried in the complex ($2,860 \pm 177 \text{ Å}^2$) greater than that buried in most antibody–antigen complexes¹⁶. Furthermore, the free energy change (ΔG) predicted on the basis of the structure for the formation of the complex (-6 kcal mol^{-1}) is in good agreement with the affinity derived from binding studies ($-11 \text{ kcal mol}^{-1}$, calculated from K_d presented in Fig. 1e), providing additional support for the physiological relevance of the crystallized complex. The interaction surface shows good shape complementarity with numerous electrostatic interactions (Fig. 2c, Supplementary Fig. 3), including multiple hydrogen bonds and salt bridges.

¹Centre for Molecular Microbiology and Infection, Imperial College, London SW7 2AZ, UK. ²Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. ³MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [†]Present address: Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

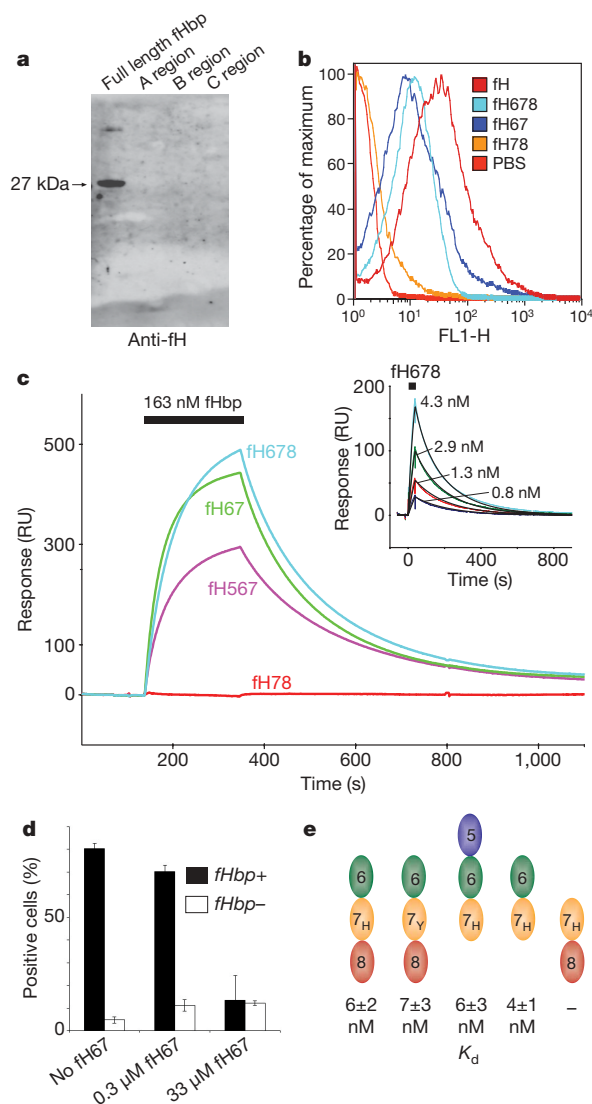


Figure 1 | The fHbp binding site requires both CCP6 of fH and requires the complete extracellular portion of fHbp. **a**, Far western analysis of fH binding to intact fHbp and truncated versions (as indicated). Membranes were incubated with purified fH which was detected with anti-fH polyclonal Abs (pAbs). Binding was only observed to the intact 27 kDa fHbp (arrowed). **b**, FACS analysis with anti-fH pAbs detected binding between *N. meningitidis* and fH constructs (all of the His form) containing CCPs 6 and 7 (note that where binding is seen the relative magnitude of the response is not significant since the amount of pAb bound is approximately proportional to the size of the fH construct rather than the tightness of the binding event). **c**, SPR demonstrates that fHbp is only able to bind fH constructs containing CCPs 6 and 7 (the number of molecules of the fH fragments coupled on the sensor channels is in the ratio 1:1:1:3 for fH78:fH567:fH678:fH67, respectively). Signals shown are after subtraction of a control trace from a mock-coupled channel. With the experiment reversed the inset shows a 1:1 Langmuir fit (black lines) to a series of fH678 dilutions (coloured lines) injected over a fHbp surface to determine kinetic parameters (e and Supplementary Table 1). **d**, FACS competition studies (using an anti-fH mAb directed against CCP5 (MRC OX24^{28,29}) and therefore unable to recognize the fH67 construct) show that the short fH67 construct (between 0.3 and 33 μ M) can compete away binding of full length fH, demonstrating that this construct contains the entire fHbp binding site. Values shown are percentage of fluorescence positive cells in a population from three experiments \pm s.d. The gate value was set using a control with no fH added. **e**, Quantitation with SPR confirms that the presence of CCPs 6 and 7 is necessary and sufficient for high affinity binding to fHbp and that the common fH polymorphism in CCP7 (402^{His/Tyr}, noted in the figure as 7_{H/7_Y}) does not significantly alter the affinity of fHbp binding.

To further probe the interaction surface, we generated doubly mutated forms of both proteins, with charged side chains being replaced by the small hydrophobic residue Ala (fH67_{R341A,H337A} and fHbp_{E283A,E304A}). Use of SPR to study the interactions between the mutated proteins and their wild-type partners revealed that the affinity of both mutant forms was reduced by more than two orders of magnitude, so that almost no interaction could be seen at analyte concentrations (~ 50 nM) around ten times the wild-type K_d (Fig. 3a). At 1,000-fold higher analyte concentration (μ M range), the mutant forms of both proteins did interact with the wild-type partners: fHbp_{E283A,E304A} retained a qualitatively similar on-rate but an increased off-rate with respect to wild-type, while fH67_{R341A,H337A} was more similar to the wild-type interaction in its off-rate but had a vastly reduced on-rate (Supplementary Table 1). It is interesting to note that E304 is not conserved in all Neisserial strains, with a Thr also found at this position¹⁰. Since these strains also sequester fH, it is likely that either the loss of a single interaction (rather than the two lost in the mutant) does not decrease the affinity sufficiently to preclude functional fH sequestration, or that subtle rearrangements of the local side chains allow the Thr side chain to interact with residues in fH. The ability of the interaction to tolerate some variation in the residues at the interface would clearly be immunologically advantageous.

A common fH polymorphism (fH 402^{His/Tyr}), recently shown to be a major susceptibility factor for age related macular degeneration^{17–19}, lies at the periphery of the interaction site, and the side chain of residue 402 (His in this structure) is within hydrogen-bonding distance of residues in fHbp (Supplementary Figs 4 and 5). However, our interaction studies do not detect any differences in binding (Fig. 1e, Supplementary Table 1) between the fH 402^{His/Tyr} isoforms. This lack of functional consequence suggests, as for variation at the 304 position (see above), two things: either (1) the loss of a single hydrogen bond from this extensive interaction surface is not sufficient to produce a functional effect; or (2) that the small amount of flexibility at the CCP 6/7 junction seen between the crystallographically independent copies of the complex allows subtle rearrangement of the fH CCP domains to enable both polymorphic variants to hydrogen bond to fHbp (Supplementary Fig. 4).

Mapping published epitopes of fHbp that elicit bactericidal antibodies^{11,12} onto the structure demonstrates that none of the sites characterized to date lie directly in the fH recognition site. Instead, epitopes recognized by antibodies that affect fH binding²⁰ are located around the edge of the recognition site and so are likely to inhibit fH binding by steric hindrance (Supplementary Fig. 5).

Our earlier structure²¹ of CCPs 6, 7 and 8 of fH in complex with sucrose octasulphate (SOS), a highly sulphated analogue of glycosaminoglycans (GAGs), revealed a charged groove on fH that forms an extended interaction site for binding of GAGs on the surface of endothelial cells. This, combined with an additional GAG-binding site in CCPs 19–20²², provides the mechanism by which fH localizes to mammalian cells²³ to prevent inappropriate complement activation. Examination of the fHbp–fH67 complex demonstrates that the Neisserial protein binding site lies within this extended GAG interaction site, overlapping the CCP6 SOS-binding site (Fig. 3b) with no fH rearrangement upon binding fHbp (Supplementary Fig. 6). There are several examples of pathogens imitating their host by expressing directly related molecules (for example, sialic acid or glycolipids²⁴). Here, however, *N. meningitidis* mimics the mechanism by which host cells regulate complement activation on their surface using amino acid side-chains instead of charged sugars. As the binding sites of fHbp and SOS on fH are overlapping, we tested whether SOS can act as a competitive inhibitor of the interaction between fH and fHbp. Strikingly, SOS prevented the fH/fHbp interaction (assessed by SPR in Fig. 3d, and by FACS in Supplementary Fig. 7) at millimolar concentrations, supporting the relevance of the crystal structure to complex formation in solution and providing a lead compound for development of small molecule inhibitors of fH/fHbp interactions.

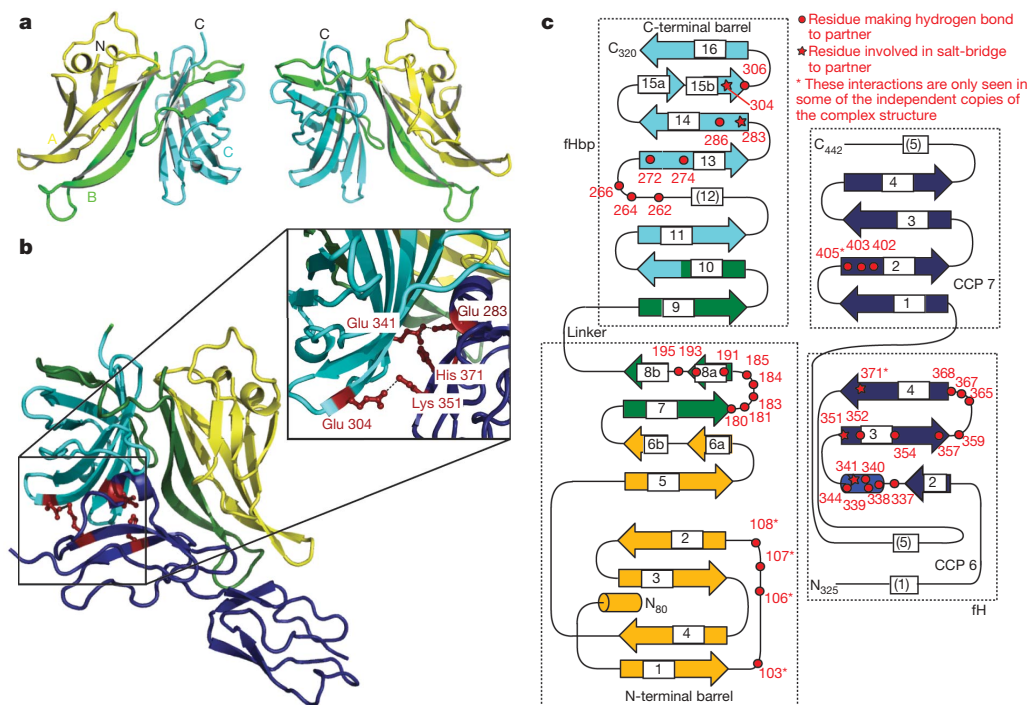


Figure 2 | Structure of fHbp and its complex with fH67. **a**, Two views of a cartoon representation of fHbp (residues 80–320). The cartoon is coloured according to the regions previously used to study fHbp (A region, yellow; B region, green; C region, cyan), illustrating the way in which these constructs do not reflect the overall architecture. **b**, Cartoon of the fHbp–fH67 complex with fHbp coloured as in **a** and fH67 in dark blue. Side chains from both proteins involved in forming salt bridges across the interaction surface are shown in red as ball-and-stick representations (zoomed and reoriented in

inset box). **c**, Topology of fHbp and fH67 coloured as in **b** with the number of the residues involved in either H-bond or salt-bridge interactions between the proteins highlighted in red. The numbering scheme used throughout is as per UniProt sequence Q9JXV4. This scheme is offset by +164 from the numbering used for the earlier NMR structure¹¹ (which was numbered from the start of their fragment) and by –65 from the scheme used in ref. 10 where numbering started from residue 1 of the mature protein without the signal sequence.

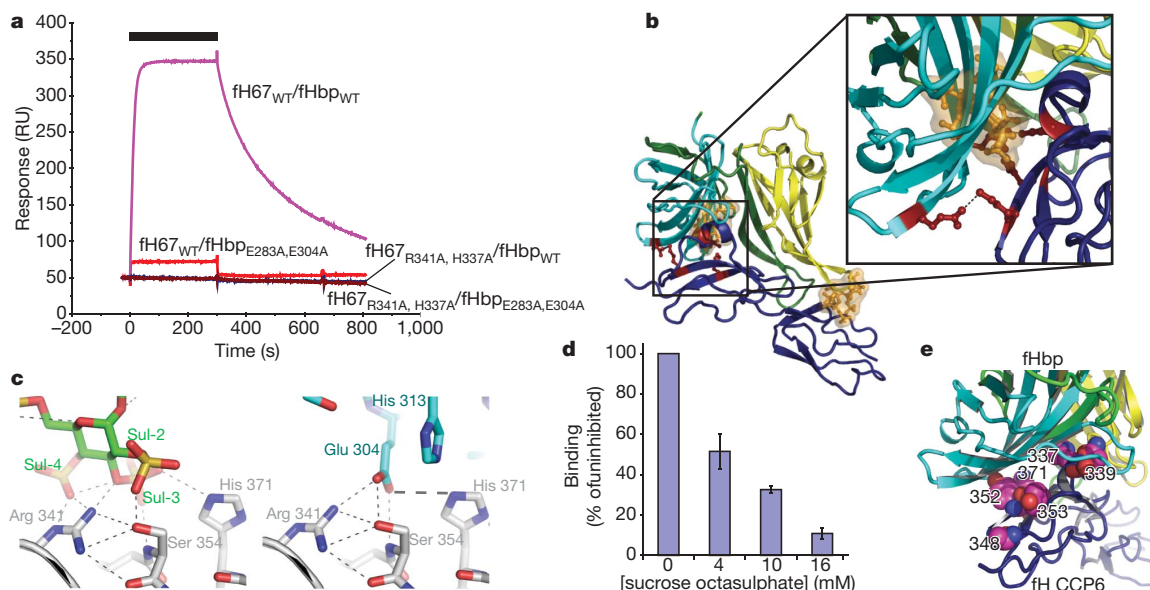


Figure 3 | Interference with fHbp binding of fH. **a**, Site-directed mutagenesis using Ala to substitute charged side-chains, shown by the structure to be involved in complex formation, abolishes binding to the wild-type form of their partner at concentrations around the wild-type K_d . The black bar indicates the time period for which the fH analytes were injected (at 50 nM, 40 $\mu\text{l min}^{-1}$) over the fHbp surfaces. **b**, Overlay of the fH-bound SOS (in gold as ball-and-stick and semi-transparent surface) from the structure of fH678 in complex with SOS²¹ onto our fHbp–fH67 structure (coloured as in Fig. 2b). This demonstrates that the sites of SOS and fHbp binding overlap, suggesting that SOS could inhibit the interaction. **c**, Mimicry of GAG

binding to fH via sulphate interactions (left panel: SOS, green carbons; fH, grey carbons) by charged side chain (Glu 304) of fHbp (right panel: fHbp, cyan carbons; fH, grey carbons). **d**, SPR demonstrates that SOS inhibits binding of fHbp to fH in the mM concentration range (fH678 injected at 4 nM, 40 $\mu\text{l min}^{-1}$ in presence of varying amounts of SOS). Values show mean \pm s.d., $n = 3$. **e**, Mapping of sites of amino acid differences in CCP6 of fH between human (UniProt P08603) and rhesus macaque (UniProt O19279). Residues which are altered shown in pink, space filling representation, with fHbp and fH67 shown as in Fig. 2b.

N. meningitidis has evolved to avoid innate responses at sites of colonization and during disease. It is also host-restricted, with the human nasopharynx its only natural reservoir⁵. This host specificity can limit the significance of animal models for studying certain aspects of pathogenesis and immunity. It has previously been shown that *N. meningitidis* only binds human fH and not fH from other primates or other animals²⁵. The amino acid sequences of fH from human and other primates are highly conserved (~90% identity) while other animal fHs are more sequence divergent. It is therefore of note that even the limited number of differences in CCP6, between human and primate sequences, cluster to the precise site of interaction with fHbp. For instance, mapping the differences between human and rhesus macaque fH²⁶ shows that all six differences within CCP6 lie in or around the interface (Fig. 3e). This indicates the likely structural basis for the species specificity of complement evasion²⁵ by *N. meningitidis* through fH recruitment, and suggests that specific modification of fH could provide a novel approach to the development of more biologically relevant models of meningococcal disease.

Interestingly, the high affinity interaction between fH and fHbp suggests that *N. meningitidis* could rapidly sequester fH in the plasma, depleting the circulating levels and de-regulating complement, rendering host cells in the vascular compartment more susceptible to complement-mediated damage. This might contribute to the dramatic haemorrhagic rash seen in meningococcal sepsis, and suggests that inhibitors specifically designed to block the interaction between CCP6 and fHbp might both promote complement-mediated clearance of bacteria and prevent sequestration of fH and consequent vascular damage²⁷.

The success of meningococcal vaccines containing fHbp that are currently being evaluated may be hampered by binding of fH which, due to the high affinity and the extensive sites of interaction, could cloak critical, conserved epitopes on the antigen. Structural analysis of this complex paves the way for the development of engineered fHbp vaccines in which fH binding is abolished to unveil the full array of protective epitopes, including those that could generate responses to inhibit recruitment of fH by this important human pathogen.

METHODS SUMMARY

Recombinant proteins were expressed and purified using affinity columns or size exclusion chromatography, and protein concentration estimated using absorbance at $\lambda = 280$ nm. Binding assays (FACS, far western and SPR) were performed according to published protocols. The complex was crystallized by vapour diffusion and X-ray diffraction data collected at the European Synchrotron Radiation Facility (ESRF, France) and DIAMOND (UK). The structure was solved by a combination of molecular replacement for the fH67 portion²¹ (one-third of the structure) and isomorphous replacement/anomalous dispersion methods, with seven independent copies refined to 2.35 Å in two different crystal forms.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 October 2008; accepted 12 January 2009.

Published online 18 February 2009.

1. Lachmann, P. J. Biological functions of the complement system. *Biochem. Soc. Trans.* **18**, 1143–1145 (1990).
2. Zipfel, P. F., Jokiranta, T. S., Hellwage, J., Koistinen, V. & Meri, S. The factor H protein family. *Immunopharmacology* **42**, 53–60 (1999).
3. Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. Complement evasion by human pathogens. *Nature Rev. Microbiol.* **6**, 132–142 (2008).
4. Jozsi, M. & Zipfel, P. F. Factor H family proteins and human diseases. *Trends Immunol.* **29**, 380–387 (2008).
5. Stephens, D. S., Greenwood, B. & Brandtzaeg, P. Epidemic meningitis, meningococcal sepsis, and *Neisseria meningitidis*. *Lancet* **369**, 2196–2210 (2007).
6. Feavers, I. M. ABC of meningococcal diversity. *Nature* **404**, 451–452 (2000).
7. Gray, S. J. et al. Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: Contribution and experiences of the Meningococcal Reference Unit. *J. Med. Microbiol.* **55**, 887–896 (2006).

8. Harrison, L. H. Prospects for vaccine prevention of meningococcal infection. *Clin. Microbiol. Rev.* **19**, 142–164 (2006).
9. Fletcher, L. D. et al. Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. *Infect. Immun.* **72**, 2088–2100 (2004).
10. Maignani, V. et al. Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. *J. Exp. Med.* **197**, 789–799 (2003).
11. Cantini, F. et al. Solution structure of the immunodominant domain of protective antigen GNA1870 of *Neisseria meningitidis*. *J. Biol. Chem.* **281**, 7220–7227 (2006).
12. Giuliani, M. M. et al. The region comprising amino acids 100 to 255 of *Neisseria meningitidis* lipoprotein GNA 1870 elicits bactericidal antibodies. *Infect. Immun.* **73**, 1151–1160 (2005).
13. Madico, G. et al. The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance. *J. Immunol.* **177**, 501–510 (2006).
14. Schneider, M. C. et al. Functional significance of factor H binding to *Neisseria meningitidis*. *J. Immunol.* **176**, 7566–7575 (2006).
15. Haralambous, E. et al. Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. *Scand. J. Infect. Dis.* **38**, 764–771 (2006).
16. Davies, D. R., Padlan, E. A. & Sheriff, S. Antibody-antigen complexes. *Annu. Rev. Biochem.* **59**, 439–473 (1990).
17. Edwards, A. O. et al. Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 (2005).
18. Hageman, G. S. et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **102**, 7227–7232 (2005).
19. Haines, J. L. et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
20. Beernink, P. T. & Granoff, D. M. Bactericidal antibody responses induced by meningococcal recombinant chimeric factor H-binding protein vaccines. *Infect. Immun.* **76**, 2568–2575 (2008).
21. Prosser, B. E. et al. Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277–2283 (2007).
22. Schmidt, C. Q. et al. A new map of glycosaminoglycan and C3b binding sites on factor H. *J. Immunol.* **181**, 2610–2619 (2008).
23. Meri, S. & Pangburn, M. K. Discrimination between activators and nonactivators of the alternative pathway of complement: Regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl Acad. Sci. USA* **87**, 3982–3986 (1990).
24. Fedtke, I., Gotz, F. & Peschel, A. Bacterial evasion of innate host defenses—the *Staphylococcus aureus* lesson. *Int. J. Med. Microbiol.* **294**, 189–194 (2004).
25. Granoff, D. M., Welsch, J. A. & Ram, S. Binding of complement factor H to *Neisseria meningitidis* is specific for human fH and inhibits complement activation by rat and rabbit sera. *Infect. Immun.* **77**, 764–769 (2009).
26. Gibbs, R. A. et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
27. Heckenberg, S. G. et al. Clinical features, outcome, and meningococcal genotype in 258 adults with meningococcal meningitis: A prospective cohort study. *Medicine (Baltimore)* **87**, 185–192 (2008).
28. Jokiranta, T. S. et al. Analysis of the recognition mechanism of the alternative pathway of complement by monoclonal anti-factor H antibodies: Evidence for multiple interactions between H and surface bound C3b. *FEBS Lett.* **393**, 297–302 (1996).
29. Sim, E., Palmer, M. S., Pukavec, M. & Sim, R. B. Monoclonal antibodies against the complement control protein factor H (beta 1 H). *Biosci. Rep.* **3**, 1119–1131 (1983).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Thanks are due to G. Bricogne and the Global Phasing Consortium for access to a β version of autoSHARP and autoBUSTER, to the beamline staff of the European Synchrotron Radiation Source (particularly G. Leonard) and DIAMOND (UK), and to many members of the Laboratory of Molecular Biophysics, Oxford, and the Lea group for help with X-ray data collection. B.E.P. was funded by a Wellcome Trust studentship; P.R. and S.J. by Medical Research Council grants to S.M.L.; J.E.D. by a Wellcome Trust grant to S.M.L.; and J.E.L. by an Engineering and Physical Sciences Research Grant to S.M.L. Work in C.M.T.'s laboratory is funded by the Wellcome Trust and the Medical Research Council. E.K. is an EMBO fellow.

Author Contributions B.E.P., J.E.D. and J.J.E.C. performed SPR experiments; B.E.P., J.J.E.C., J.E.L. and S.Q. expressed proteins and crystallized the complex; M.C.S., E.K. and Q.Z. performed all assays with *Neisseria* and expressed fHbp; S.J., P.R. and S.M.L. collected and analysed X ray data; S.M.L. phased the X-ray data and built/refined the models; and S.M.L. and C.M.T. designed the research and wrote the paper.

Author Information Coordinates and X-ray data have been deposited in the Protein Data Bank under accession numbers 2w80 and 2w81. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M.L. (susan.lea@path.ox.ac.uk) or C.M.T. (c.tang@imperial.ac.uk).

METHODS

Western blot analysis. Protein samples were separated by SDS–PAGE, then transferred to PVDF membranes at 15 V for 60 min. The membranes were blocked in 5% milk in PBS at 4 °C overnight, rinsed once in PBST (PBS, 0.05% Tween20, 0.5% skimmed milk), then incubated with primary antibodies or a source of fH for 1–2 h at room temperature. After three washes in PBST, membranes were incubated with goat anti-human fH pAbs (at a 1:1,000 final dilution) or murine anti-fHbp pAbs (at a 1:10,000 final dilution) in PBS. Binding was detected with an anti-rabbit peroxidase conjugated IgG (Dakocytomation) used at 1:500 dilution.

Flow cytometry analysis. A total of 10^8 *N. meningitidis* were incubated with 10 μ l of purified fH (2 μ M) or CCPs (final concentration, 3 μ M) for 30 min at 37 °C. After three washes with PBS–0.1% Tween20 (PBST), bacteria were incubated for 30 min with a mAb MRC OX24^{28,29} or with anti-human fH pAbs in PBS. After three washes in PBST, cells were incubated with a secondary, FITC-conjugated antibody for 30 min in the dark at 4 °C. fH bound to meningococci was detected by flow cytometry using a FACS Calibur analyser (Becton Dickinson). For inhibition assays, full length fH was incubated with bacteria together with increasing concentrations of CCPs 6 and 7 as indicated with binding of full length fH detected using the anti-CCP5 mAb MRC OX24^{28,29}, or with SOS (Toronto Research Chemicals). Results are expressed as the mean fluorescence index (geometric mean fluorescence multiplied by the percentage positive cells).

Expression of proteins and purification of the complex. fHbp and truncated versions of the protein were expressed in *E. coli* BL21(DE3) host cells as poly-histidine-tagged fusion proteins at their C terminus¹⁴. Cultures of the strains were grown to mid-log phase and expression induced with isopropyl-D-thiogalactoside (IPTG, 1 mM final concentration). Proteins were purified by affinity chromatography on a His-Trap column according to the manufacturer's instructions (GE Healthcare), and eluted from columns with 250 mM imidazole. fH constructs were expressed in *E. coli* as previously described³⁰ (with refolding to allow correct formation of the four disulphide bonds within the two CCP fragments) and purified either by heparin column affinity purification or by size exclusion chromatography. Following mixing of fHbp with fH67 (at a large molar excess), the complex was purified by size exclusion chromatography (Superdex S-200, GE Healthcare, buffer 50 mM Tris pH 7.5, 150 mM NaCl) where the strong interaction seen between the uncomplexed fH67 and the column matrix ensured good separation of the peak containing the complex and that containing the excess fH67. Presence of both proteins in the putative complex peak was confirmed by SDS–PAGE and mass spectrometry (data not shown).

Site-directed mutagenesis. Double mutants were generated in a sequential fashion using the QuikChange site-directed mutagenesis kit (Stratagene). The R341A, H337A double mutant form of fH67 was generated using the following primers: H337A forward, 5'-AAACATGGAGGCTCTATGCTGAGAATATGCGTAGACCATACTTTCC-3'; H337A reverse, 5'-TTTGTACCTCCAGATACGACTCTTATACGCATCTGGTATGAAAGG-3'. R341A forward, 5'-GGTCTATATCATGAGAATATGGCTAGACCATACTTTCCAGTAGC-3'; R341A reverse, 5'-CCAGATATAGTACTCTTATACCGATCTGGTATGAAAGGTCATCG-3'. Introduction of the mutation was confirmed by sequencing. Recombinant mutant fH67 fragments were expressed and refolded from inclusion bodies as previously described³⁰. For generation of the E283A, E304A fHbp the primers used were: E283A forward 5'-ACAACCAAGCCGCGAAAGGCAGTTAC-3'; E283A reverse, 5'-GTAACCTGCCTTTGCGGGCTTGGTTGT-3'; E304A forward 5'-TGCCGGCAGCGCGGCAGTGAACCG-3'; E304A reverse 5'-CGGTTTTCATCGCCGCGCTGCCGCA-3' and the protein expressed and purified as described above.

Surface plasmon resonance. Either fHbp or fH constructs were coupled to the sensor surface (CM5 chips) using standard amine coupling protocols. The appropriate ligand was then flowed in the fluid phase (buffer HEPES buffered saline) at a flow rate of 40 μ l min⁻¹. A control trace was collected using either a mock-coupled channel or an unrelated protein-coupled channel. Fits were performed to control-subtracted traces such as those shown in Fig. 1. The strong interaction was not disrupted using either high/low salt (0 to 3 M NaCl) or extreme pH (range 4 to 8 tried) and extended dissociation time (60 min) was therefore used between successive injections. The data obtained with the larger fHbp in the fluid phase could not be satisfactorily fitted using a simple 1:1 model (presumably due either to issues of movement of the larger protein into/out of the chip matrix or steric hindrance in the fH binding site introduced by coupling of the smaller protein to the chip surface), crude analysis of near-equilibrium values suggested a K_d in the tens to hundreds of nM range (depending on the fH construct used). Once the interaction was reversed (with fH on the chip) the data could be satisfactorily fitted (Supplementary Table 1; Fig. 1c, inset and Supplementary Fig. 1) using a 1:1 Langmuir model with satisfactory χ^2 revealing that the true K_d for the constructs containing CCP6 was of the order of 5 nM irrespective of the flow rate, injection time and density of fHbp coupled to the

chip surface (see main text, Supplementary Table 1). The fH67 and fHbp double mutants were also analysed with the fHbp partner coupled to the chip surface (fHbp_{WT} ~800 RU coupled on channel 1, fHbp_{E283A,E304A} ~1,500 Response Units (RU) coupled on channel 2). fH67 (wild type and mutant) were then flowed over at a concentration of 50 μ M (~10 \times the native K_d) revealing little or no interaction from the mutants with their wild type partner compared to the magnitude of interaction seen with both proteins in the wild type form. To calculate the affinity of fH67_{R341A,H337A} for fHbp_{WT}, a series of injections (in duplicate) spanning the concentration range 250 nM to 20 μ M was used and fitted using the BiaEvaluation software to give the values shown in Supplementary Table 1. Inhibition assays were calculated by injecting a constant concentration of fH67 or fH678 over a fHbp-coupled surface with/without inhibitor (SOS in Fig. 3d or heparin in Supplementary Fig. 8) present. The binding signal obtained is then reported as a percentage of the binding seen in the absence of inhibitor. Values are given as a mean with standard deviations as shown from three independent pairs of injections. SPR studies using full length fH purified from plasma also showed an interaction (Supplementary Fig. 9) with fHbp on the sensor chip surface and fH in the fluid phase. These data could not be satisfactorily fitted, presumably due to complex kinetics caused by the difficulties of diffusing a large, flexible, molecule into the dextran matrix. However, concentrations of fH in the 10 nM range could be seen to interact, suggesting that the affinity for the full length fH is in the same range as that for the smaller fragments. Attempts to couple full length fH to the chip surface resulted in a coupling which prevented binding of fHbp (data not shown) and so could not be used to quantify the interaction.

Crystallization and X-ray data collection. The crystals were grown using the sitting drop vapour diffusion method from 0.2 μ l complex + 0.2 μ l mother liquor drops (complex at 4.5 mg ml⁻¹ calculated using an assumed extinction coefficient of 2.2, mother liquor optimized around JCSG-plus condition 15, 20% polyethylene glycol (molecular weight 6,000), 0.1 M Bicine pH 9.0). Harvesting of crystals and SDS–PAGE confirmed that both fH67 and fHbp were present (data not shown). Crystals (usually 50 \times 10 \times 5 μ m) were flash frozen in liquid nitrogen following cryoprotection with 15% ethylene glycol. Data were collected at both the ESRF and DIAMOND by the rotation method using either an oscillation range of 0.5° or 1° with the crystal kept at 120 K. The native P1 data were collected at beamline ID14eh4 (ESRF) with λ = 0.9523 Å, the native C2 and Pt sets were collected at beamline I03 (DIAMOND) with λ = 0.9814 Å, the Hg set at beamline BM14 (ESRF) with λ = 1.003 Å and the S-SAD at beamline ID29 (ESRF) with λ = 1.8 Å. Data were integrated and scaled using xia2 (<http://www.ccp4.ac.uk/xia/>)³¹ with the -3d option to force use of program XDS³² for integration and program Scala³³ for inter-frame scaling.

Structure solution and refinement. The structure of the fH67–fHbp complex was solved in the C2 crystal form by a combination of molecular replacement with the structure of fH67 (taken from the earlier structure of fH678 in complex with sucrose octasulphate, SOS, PDB ID 2UWN²¹) and MIRAS methods using Pt, Hg derivatives and the anomalous scattering from 31 of the 33 sulphurs present in the C2 asymmetric unit. Three copies of fH67 were found using MolRep³⁵ (CCP4 Program Suite³⁶) and the two CCP domains rigid body refined using Buster-TNT³⁷ without modelling of missing atoms. These phases were then incorporated into a SHARP³⁸ phasing job and used to locate the heavy atom sites in the other crystals (6 Pt, 1 Hg and 31 S). The heavy atom model was refined in SHARP³⁸ (with B factor fixed to the Wilson value for each crystal) using data to 3.2 Å and the phases generated combined with those from the molecular replacement within SHARP to generate phases with a figure of merit of 0.54 (0.35 for the highest resolution shell 3.3–3.2 Å). The phasing powers for the derivatives were (anomalous/isomorphous) 0.7/2.4 for the Pt, 0.2/0.3 for the Hg and 0.3/0.0 for the anomalous sulphur signal. The map obtained using these phases was solvent flattened within SHARP³⁸ and averaged in DM³⁹ using three separate operators for the fHbp, fH CCP6 and fH CCP7 regions. The NMR model¹¹ for the C-terminal barrel of fHbp could not be placed in this density (the strands were seen to be distorted) so a model was built by hand using program Coot⁴⁰. Rebuilding and refinement using data in the range 40–2.35 Å proceeded using Coot⁴⁰ and Buster-TNT³⁷ with non-crystallographic symmetry used to restrain the core of each domain while more variation was allowed between the loop region until the R/R_{free} were 25.5/27.1 with no residues in the disallowed regions of the Ramachandran plot (94.6% in the most favoured, as determined by MolProbity⁴¹). The model contains all residues of fH67 and all but five residues at the N terminus of fHbp and the six histidines of the affinity tag which are disordered in all copies of the molecule. Molecular replacement of the individual domains (fH CCP6 and CCP7 and fHbp) using program MolRep³⁵ allowed solution of the P1 crystal form (four independent copies of the complex in a different packing arrangement). Minor rebuilding of the loop regions and refinement (90–2.35 Å) using Coot⁴⁰/Buster-TNT³⁷ led to R/R_{free} of 23.2/26.3 once again with no residues in the disallowed regions of the Ramachandran plot

(95.3% in the most favoured). Structural figures are drawn using program PyMol (<http://www.pymol.org>).

30. Prosser, B. E. *et al.* Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulfate/human complement regulator factor H SCRs 6-8. *Acta Crystallogr. F* **63**, 480–483 (2007).
31. Bahar, M. *et al.* SPINE workshop on automated X-ray analysis: a progress report. *Acta Crystallogr. D* **62**, 1170–1183 (2006).
32. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
33. Evans, P. R. Data reduction. In *Proceedings of CCP4 Study Weekend* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 114–122 (SERC, Daresbury Laboratory, 1993).
34. Prosser, B. E. *et al.* Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277–2283 (2007).
35. Vagin, A. A. & Isupov, M. N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D* **57**, 1451–1456 (2001).
36. Winn, M. D. An overview of the CCP4 project in protein crystallography: An example of a collaborative project. *J. Synchrotron Radiat.* **10**, 23–25 (2003).
37. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
38. De La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength-anomalous dispersion methods. *Methods Enzymol.* **276**, 472–494 (1997).
39. Cowtan, K. & Main, P. Miscellaneous algorithms for density modification. *Acta Crystallogr. D* **54**, 487–493 (1998).
40. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
41. Davis, I. W. *et al.* MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–383 (2007).

LETTERS

Inactivation of the *Fto* gene protects from obesity

Julia Fischer^{1*}, Linda Koch^{2,3,4*}, Christian Emmerling¹, Jeanette Vierkotten^{1†}, Thomas Peters^{1†}, Jens C. Brüning^{2,3,4} & Ulrich Rüther¹

Several independent, genome-wide association studies have identified a strong correlation between body mass index and polymorphisms in the human *FTO* gene^{1–4}. Common variants in the first intron define a risk allele predisposing to obesity, with homozygotes for the risk allele weighing approximately 3 kilograms more than homozygotes for the low risk allele¹. Nevertheless, the functional role of *FTO* in energy homeostasis remains elusive. Here we show that the loss of *Fto* in mice leads to postnatal growth retardation and a significant reduction in adipose tissue and lean body mass. The leanness of *Fto*-deficient mice develops as a consequence of increased energy expenditure and systemic sympathetic activation, despite decreased spontaneous locomotor activity and relative hyperphagia. Taken together, these experiments provide, to our knowledge, the first direct demonstration that *Fto* is functionally involved in energy homeostasis by the control of energy expenditure.

Previously, we have identified the murine homologue *Fto* as part of a 1.6-megabase (Mb) deletion on chromosome 8 in the *Fused toes* mouse mutant, a mutant with severe developmental defects^{5–7}. *Fto* contains a nuclear localization signal and shows a ubiquitous expression pattern^{1,5,8}. Computational analysis suggests that *Fto* belongs to the FeII- and 2-oxoglutarate-dependent dioxygenase superfamily^{8,9}. *In vitro* analyses show that it catalyses the demethylation of 3-methylthymine in single-stranded DNA⁸. To directly investigate the role of *Fto* in energy homeostasis *in vivo*, we have generated *Fto*-negative (*Fto*^{−/−}) mice and have characterized them with respect to energy and glucose homeostasis.

Fto-deficient mice were generated by replacing exons 2 and 3 of the *Fto* gene with a neomycin resistance cassette (Supplementary Fig. 1a, b). *Fto*^{−/−} mice were born at the expected Mendelian ratio, suggesting that *Fto* is not essential for embryonic development. Homozygous mutant mice are viable but postnatal death occurs more frequently, resulting in a ratio of 1:1.94:0.61 (*Fto*^{+/+}:*Fto*^{+/-}:*Fto*^{−/−}) at 4 weeks of age. We speculate that the reduced viability of *Fto*^{−/−} mice is due to phenotypic disadvantages as the mortality rate is decreased by postponed weaning and a reduction in litter size.

Successful deletion of *Fto*, resulting in the absence of the *Fto* protein, was shown by western blot analysis of different tissues, such as brain, liver and skeletal muscle (Fig. 1a). Immunofluorescence analysis of different tissues further revealed the nuclear localization of the endogenous protein in control mice and confirmed the absence of the *Fto* protein in *Fto*^{−/−} mice (Fig. 1b and Supplementary Fig. 1c).

The *Ftm* (also known as *Rpgrip11*) gene lies, in opposite orientation, in close proximity to *Fto* and it has therefore been considered as a further candidate gene for the association with obesity^{1,10}. To rule out the possibility that the expression of *Ftm* is affected by modification of the *Fto* locus in the mutant mice, we compared *Ftm* expression

in various metabolically relevant tissues of *Fto*^{−/−} and *Fto*^{+/+} mice. PCR with reverse transcription (RT-PCR) analysis showed that *Ftm* expression was unaltered, thereby excluding the possibility that any phenotype observed in these mice is based on the deletion of regulatory elements of the *Ftm* gene (Fig. 1c).

The effects of *Fto* deficiency manifest shortly after birth as *Fto*^{−/−} mice have unaltered body weight during embryonic development and at birth, but show a growth retardation from postnatal day 2 onwards (Fig. 2a). We monitored body weight over a period of several weeks and detected an average weight reduction of approximately 30–40% (males and females) after 6 weeks, whereas the body weight of heterozygous mice does not significantly differ from *Fto*^{+/+} mice (Fig. 2b, c). The anatomical characteristics of *Fto*-deficient mice, however, appeared unchanged (Fig. 2d). Although the differences between wild-type and homozygous animals are obvious at all examined ages, the heterozygotes show more subtle changes. At the age of 20 weeks, female *Fto*^{+/-} mice show a trend towards reduced body weight, whereas heterozygous males do not differ from wild-type littermates (Fig. 2e). As well as a significant reduction in body weight, the nose–anus length of *Fto*^{−/−} mice is reduced compared to wild-type animals from 10.1 to 8.6 cm, and 10.3 to 9.4 cm, in males and females, respectively (Fig. 2f). The reduced size and body weight of the *Fto*^{−/−} mice persisted throughout their entire lifespan (Supplementary Fig. 2a, b), supporting the notion that *Fto* is involved in processes connected to metabolism and growth regulation.

Considering the suggested function of *Fto* in the control of energy homeostasis and the development of obesity, we next analysed the body composition of control and *Fto*^{−/−} mice. Magnetic resonance imaging (MRI) analysis showed that fat mass is significantly reduced in *Fto*-negative mice by 60% and 23% in males and females, respectively (Fig. 3a). Although lean mass is also reduced in these mice, it is affected to a lesser extent than fat mass, namely by 26% and 19% in males and females (Fig. 3b). Consistently, we observed a significant reduction in the mass of dissected epididymal white adipose tissue (WAT), which became more pronounced with increasing age, resulting in a nearly complete loss of WAT at the age of 15 months (Fig. 3c and Supplementary Fig. 2b, c). Haematoxylin and eosin staining showed a markedly reduced adipocyte size in the *Fto*^{−/−} mice (Fig. 3d). Conversely, analysis of brown adipose tissue (BAT) mass showed no significant changes (Fig. 3e).

Given the reduced body length and obesity of *Fto*-deficient mice, we next investigated the effect of partially or completely reduced *Fto* expression on body weight and adipose tissue after exposure to a high fat diet (HFD). After 12 weeks on a HFD, male *Fto*^{−/−} and *Fto*^{+/-} mice showed a significantly reduced weight gain compared to *Fto*^{+/+} mice (Fig. 3f). Concomitantly, both genotypes accumulated significantly less WAT (Fig. 3g), indicating that already a moderate reduction in

¹Institute for Animal Developmental and Molecular Biology, Heinrich Heine University, Universitätsstr. 1, D-40225 Düsseldorf, Germany. ²Department of Mouse Genetics and Metabolism, Institute for Genetics and Second Department of Internal Medicine, ³Center of Molecular Medicine Cologne (CMMC) and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, D-50674 Cologne, Germany. ⁴Max-Planck Institute for the Biology of Ageing, D-50674 Cologne, Germany. [†]Present addresses: Department of Dermatology, Center of Molecular Medicine Cologne (CMMC), University of Cologne, D-50674 Cologne, Germany (J.V.); Epidauros Biotechnologie AG, D-82347 Bernried, Germany (T.P.).

*These authors contributed equally to this work.

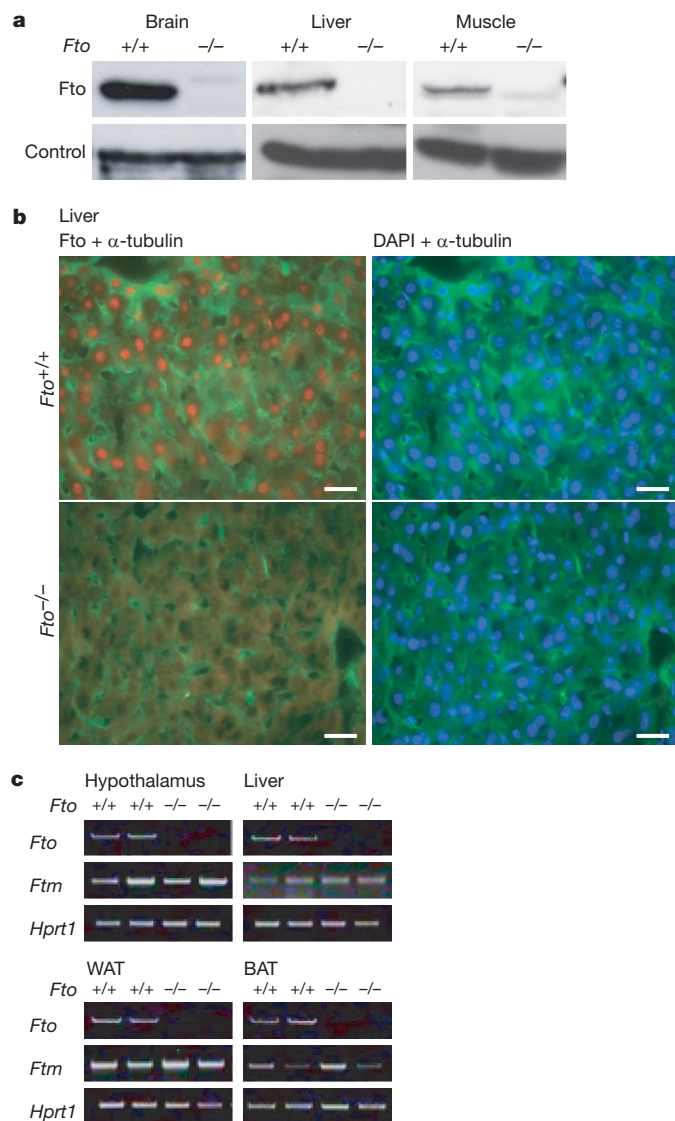


Figure 1 | Successful generation of *Fto*-deficient mice. **a**, Western blot analysis with an *Fto*-specific antibody shows a 58-kDa protein that is not present in *Fto*^{-/-} tissues, confirming the successful deletion. Loading control was α -tubulin for brain and muscle, β -actin for liver. **b**, Immunofluorescence analyses with the *Fto*-specific antibody show nuclear localization of the endogenous protein that cannot be found in *Fto*^{-/-} cells. Counterstaining was performed with α -tubulin antibody and 4,6-diamidino-2-phenylindole (DAPI). Scale bars, 25 μ m. **c**, RT-PCR analysis of *Fto*^{+/+} and *Fto*^{-/-} tissues. Although *Fto* mRNA cannot be detected in *Fto*^{-/-} tissues, *Ftm* expression is unchanged. The *Hprt1* house-keeping gene was used as a positive control.

Fto expression, as present in the heterozygotes, protects from diet-induced obesity.

Owing to the marked reduction of adipose tissue in *Fto*^{-/-} mice, we proposed that changes in the secretion of adipokines occur, which are highly correlated to metabolic disorders, weight regulation and body fat^{11–18}. As expected, serum leptin concentrations were significantly decreased in *Fto*^{-/-} mice corresponding to the reduced adipose tissue mass seen in these animals (Supplementary Fig. 2d). Accordingly, serum concentrations of adiponectin, which has been shown to be inversely correlated to adipose tissue mass, were significantly increased in *Fto*^{-/-} mice (Supplementary Fig. 2e)¹⁸.

As adiposity is determined by the disbalance between food intake and caloric utilization, we next investigated these parameters of energy homeostasis in control and *Fto*^{-/-} mice. This analysis demonstrated that the absolute food intake of *Fto*^{-/-} mice was unaltered compared

to littermate controls (Fig. 4a). When correcting the data either for body weight (data not shown) or for lean mass, however, *Fto*^{-/-} mice show a significant relative hyperphagia (Fig. 4b). These experiments indicate that the lean phenotype of *Fto*^{-/-} mice does not result from a reduced caloric intake.

Because *Fto* has been previously demonstrated to be expressed particularly in the arcuate nucleus (ARC) of the hypothalamus, a well-defined key regulatory centre of energy homeostasis, and moreover to undergo changes in messenger RNA expression depending on the energy state^{8,10,19}, we next examined whether loss of *Fto* expression results in gross alterations of hypothalamic development or maintenance. The earliest steps of hypothalamus formation are regulated by Sonic hedgehog (Shh), which induces hypothalamic gene expression²⁰. Normal expression of Shh at embryonic day (E) 10.5 in *Fto*^{-/-} mice, particularly in the diencephalon, indicates an unaltered induction of hypothalamus development (Supplementary Fig. 3a). Furthermore, the neuroanatomy of the hypothalamus of adult mice was unchanged, ruling out major alterations of hypothalamic development as a cause for the phenotype of *Fto*-deficient mice (Supplementary Fig. 3b).

Because control of the melanocortin system comprising pro-opiomelanocortin- α (Pomc)-expressing and agouti-related peptide (AgRP)/neuropeptide Y (NPY)-expressing neurons in the ARC represents a pivotal regulatory mechanism in energy homeostasis, we next investigated the expression of *Fto* in Pomc neurons in brain sections of mice, which express green fluorescent protein (GFP) under the control of the *Pomc* promoter. This co-localization study clearly confirmed abundant expression of *Fto* in the ARC, which overlaps with, but is not restricted to Pomc neurons (Supplementary Fig. 3c). To address further whether the loss of *Fto* expression affects regulation of neuropeptide expression, we performed mRNA expression analyses in the fed and fasted state. This analysis revealed that expression of these neuropeptides did not significantly differ between control and *Fto*-deficient mice in the fed state (Fig. 4c). However, whereas *AgRP* mRNA expression increased to a comparable extent in control and *Fto*^{-/-} mice after fasting, starvation-induced *Npy* mRNA induction was blunted and starvation-induced *Pomc* mRNA repression was slightly exaggerated in *Fto*-deficient mice (Fig. 4c and Supplementary Fig. 3d, e). However, regulation of *Npy* and *Pomc* mRNA expression remained unchanged in heterozygous *Fto*^{+/-} mice (Supplementary Fig. 3d, e). Taken together, *Fto*-deficiency has no significant effect on neuropeptide expression in the fed state, but slightly reduces both *Npy* and *Pomc* expression in the fasted state, consistent with the lack of hypophagia in the *Fto*-deficient mice. Moreover, normal basal neuropeptide expression provides further support for normal development and maintenance of the AgRP/NPY and Pomc neurons in the ARC of the hypothalamus.

To assess whether increased energy expenditure accounts for the reduced adiposity of *Fto*^{-/-} mice in the presence of an increased relative caloric uptake, we next determined the energy expenditure of the animals using indirect calorimetric analysis. The oxygen consumption, carbon dioxide production and calculated heat generation of *Fto*^{-/-} mice were significantly increased both during the day and night period, indicating a higher energy expenditure of *Fto*^{-/-} mice compared to their *Fto*^{+/+} littermates (Fig. 4d, e and Supplementary Fig. 4a). These data indicate that the leanness of *Fto*-deficient mice develops as a consequence of increased energy expenditure.

To analyse further the potential mechanisms underlying increased energy expenditure of *Fto*-deficient mice, we next measured the physical activity of the animals. This analysis showed that *Fto*-negative mice had significantly decreased physical activity, and a similar trend could be observed in the heterozygous animals (Fig. 4f and Supplementary Fig. 4b). Thus, *Fto*-deficient mice have increased energy expenditure in the presence of reduced spontaneous locomotor activity. Reduced locomotor activity occurred in the presence of unaltered MCH expression, as a known regulator of locomotor activity²¹ (Supplementary Fig. 4c).

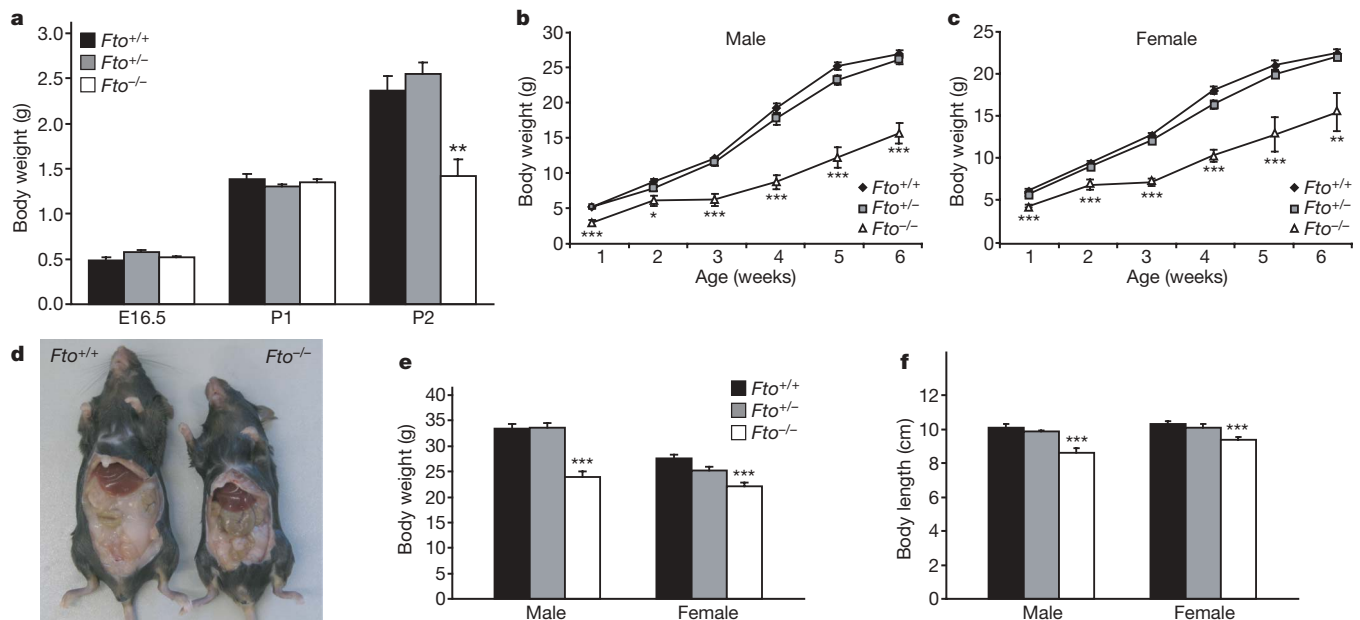


Figure 2 | Phenotypic characteristics of *Fto*-negative mice. **a**, Embryos and animals at E16.5 and on postnatal day (P)1 and P2 were weighed. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, E16.5: *n* = 6/10/5, P1: *n* = 6/7/5, P2: *n* = 12/22/7; $^{**}P = 0.0025$). **b**, **c**, Body weight of mice of different genotypes and gender shows a reduced body weight of *Fto*^{-/-} mice. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, males: *n* = 7/12/3, *P*-values from week 1–6: 2.7×10^{-4} , 0.02, 4.2×10^{-5} , 1×10^{-4} , 1.4×10^{-5} and 4.2×10^{-5} ; females: *n* = 9/18/3, *P*-values from week 1–6: 0.001, 0.001, 1.7×10^{-7} , 1.8×10^{-5} , 6×10^{-4} and 0.002). **d**, Anatomical

view of *Fto*^{+/+} and *Fto*^{-/-} mice aged 20 weeks shows the reduced body length and leanness of *Fto*^{-/-} mice. **e**, Reduced body weight of *Fto*^{-/-} mice aged 20 weeks. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, males: *n* = 15/17/15, $^{***}P = 3.1 \times 10^{-6}$; females: *n* = 15/18/21, $^{***}P = 0.0001$). **f**, Body length of 20-week-old mice. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, males: *n* = 11/10/8, $^{***}P = 0.0004$; females: *n* = 14/8/20, $^{***}P = 9.1 \times 10^{-5}$). All values are mean and s.e.m.

Moreover, neither increased mitochondrial uncoupling nor altered thyroid function seemed to account for the increased energy expenditure of *Fto*^{-/-} mice, because analysis of circulating thyroid hormone concentrations as well as the mRNA expression of uncoupling protein *Ucp1* in BAT revealed no changes in the absence of functional *Fto* expression (Supplementary Fig. 4d, e). In contrast, circulating plasma adrenaline concentrations were significantly

increased in *Fto*^{-/-} compared to control mice, and circulating plasma noradrenaline concentrations showed a similar trend, which did not reach statistical significance (Fig. 4g, h). Thus, energy expenditure in *Fto*-deficient mice occurs in the presence of increased sympathetic activation.

To investigate whether the leanness and increase in energy expenditure of *Fto*^{-/-} mice leads to alterations in glucose metabolism, we

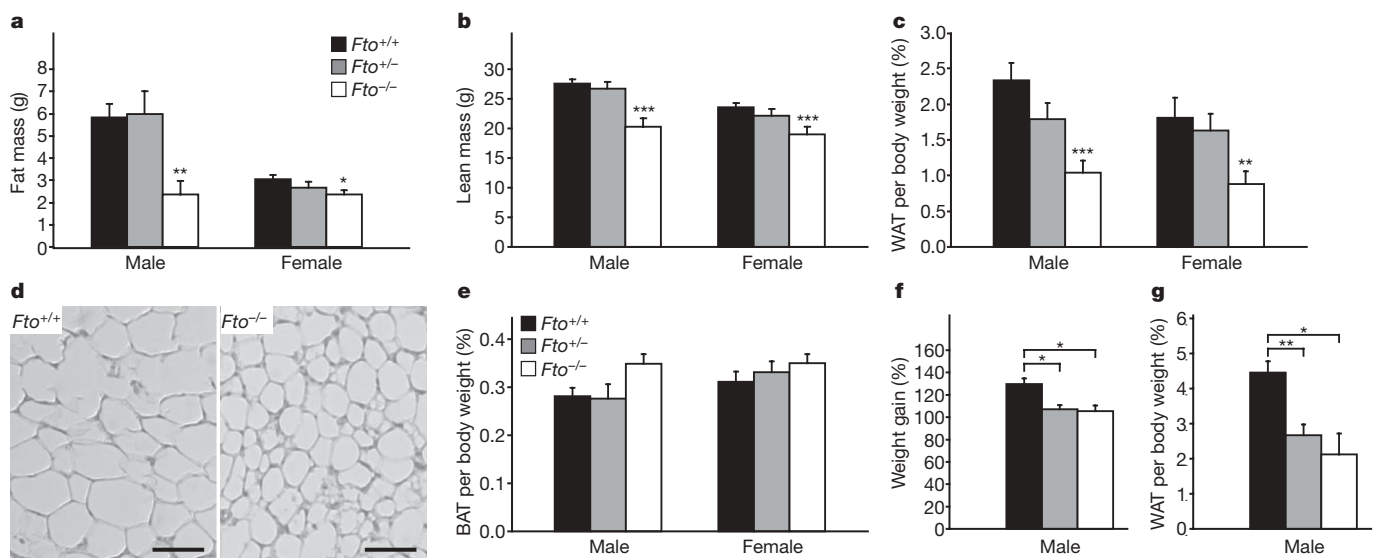


Figure 3 | Body composition of *Fto*-negative mice. **a**, **b**, MRI analysis showed that fat mass, and to a lesser extent lean mass, is significantly reduced in *Fto*-negative animals. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, males: *n* = 9/4/4, $^{**}P = 0.0078$, $^{***}P = 0.0006$; females: *n* = 12/14/19, $^{*}P = 0.0228$, $^{***}P = 1.4 \times 10^{-5}$). **c–e**, WAT (**c**) and BAT (**e**) was isolated from 20-week-old animals, weighed and correlated to body weight. Haematoxylin and eosin staining of WAT showed a prominent reduction of adipocyte size in *Fto*^{-/-} mice (**d**). Scale bars, 50 μ m. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, males: *n* = 12/

11/15, $^{***}P = 0.0004$ (**c**), *n* = 12/11/11 (**e**); females: *n* = 15/18/21 (**c**, **e**), $^{**}P = 0.0078$). **f**, **g**, Male *Fto*^{+/+} and *Fto*^{+/-} mice of similar weight aged 8 weeks (*Fto*^{+/+}: 28.30 g, *Fto*^{+/-}: 28.65 g) and male *Fto*^{-/-} mice were exposed to a HFD for 12 weeks. The weight gain was calculated in correlation to control mice (**f**). Epigonadal WAT was isolated, weighed and correlated to body weight (**g**). (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, *n* = 5/8/3, $^{*}P = 0.042$ (*Fto*^{+/+}), $^{*}P = 0.033$ (*Fto*^{-/-}) (**f**), *n* = 5/7/3, $^{**}P = 0.0012$ (*Fto*^{+/+}), $^{*}P = 0.018$ (*Fto*^{-/-}) (**g**). All values are mean and s.e.m.

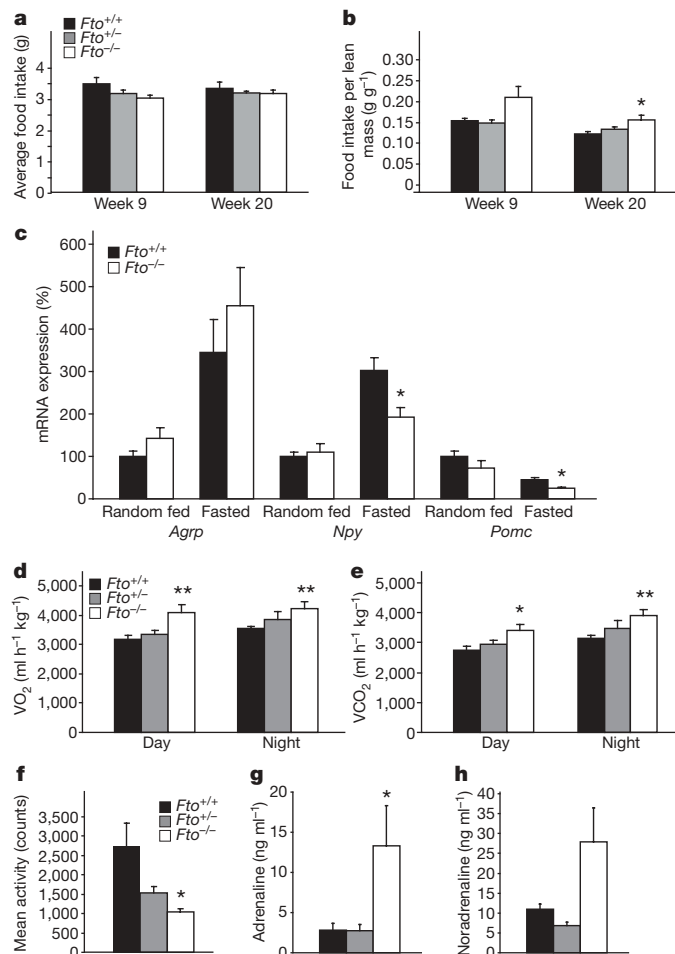


Figure 4 | Parameters of energy uptake and expenditure. **a, b,** Food intake of female mice is depicted as the average amount of food consumed over a period of 24 h. The total food intake of *Fto*^{-/-} mice was unaltered (**a**), but when correlated to lean mass, *Fto*^{-/-} mice showed an increase in food intake (**b**). (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, week 9: *n* = 8/5/9, week 20: *n* = 9/9/13; **P* = 0.015). **c,** Relative expression of neuropeptides in the hypothalamus was determined by RT-PCR. (*Fto*^{+/+}/*Fto*^{+/-}: random fed: *n* = 5/7, fasted: *n* = 14/14, **P* = 0.03 (*Npy*), **P* = 0.04 (*Pomc*)). **d, e,** The volume of O₂ consumed (VO₂) (**d**) and of CO₂ produced (VCO₂) (**e**) were measured over 24 h and correlated to lean mass. Significantly higher energy expenditure in *Fto*^{-/-} mice was detected. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, *n* = 10/5/8, ***P* = 0.0057 (VO₂, day), **P* = 0.0130 (VCO₂, day), ***P* = 0.0072 (VO₂, night), ***P* = 0.0034 (VCO₂, night)). **f,** Physical activity was measured for 60 h and is depicted as the average total physical activity. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, *n* = 6/5/5, **P* = 0.039). **g, h,** Adrenaline (**g**) and noradrenaline (**h**) levels were determined by ELISA. (*Fto*^{+/+}/*Fto*^{+/-}/*Fto*^{-/-}, *n* = 7/6/5, **P* = 0.0490 (**g**); *n* = 7/6/6 (**h**)). All values are mean and s.e.m.

compared different parameters of glucose metabolism between *Fto*^{-/-} mice and controls. We found an overall unaltered glucose metabolism with a mild improvement in insulin sensitivity, presumably as a consequence of leanness and increased circulating plasma adiponectin concentrations (Supplementary Fig. 5a–g).

Human *FTO* has been identified as a candidate gene predisposing to obesity by a correlation of variants in intron 1 to body mass index^{1–4}. Until now, however, it is unclear whether these variants have any influence on *FTO* regulation. Here we demonstrate that loss of the mouse *Fto* gene product reduces adiposity, and that even moderate reduction in *Fto* expression protects from diet-induced obesity. Our data suggests that variants in the human *FTO* gene might lead to an up- or dysregulation of *FTO* expression, thus rendering individuals more susceptible to obesity. Interestingly, whereas carriers of the human *FTO* risk allele seem to develop obesity as a consequence of

hyperphagia apparently in the absence of altered energy expenditure^{22–25}, leanness in the absence of *Fto* in mice results from increased energy expenditure in the presence of relative hyperphagia, and is associated with increased sympathetic tone. Although sympathetic activation has been demonstrated to control energy expenditure partly by the regulation of mitochondrial uncoupling, it can also activate substrate cycling of fatty acids/triglycerides as an alternative mechanism of energy expenditure^{26–28}. Thus, in the presence of unaltered *Ucp1* expression in *Fto*-deficient mice, futile cycling represents a probable cause of increased energy expenditure. Further studies will have to determine the molecular basis of *FTO*-controlled sympathetic regulation and potential direct mechanisms in control of substrate cycling, possibly leading to the identification of new therapeutic targets in the treatment of obesity.

METHODS SUMMARY

Detailed methods are included in the Methods. The *Fto* gene was inactivated by the targeted deletion of exons 2 and 3, as illustrated in the Supplementary Fig. 1. All animal procedures and euthanasia were reviewed by the animal care committee, approved by local government authorities and were in accordance with National Institutes of Health guidelines. Blood analysis, extraction and detection of mRNA and proteins and immunofluorescence analysis were performed following standard protocols. Details of antibodies, primers and probes are given in the Methods. Whole body composition of live animals was determined using the NMR Analyser Minispec (Bruker Optik). Indirect calorimetry was measured in a Calorimetry Module (CaloSys V2.1, TSE Systems) for 24 h. Locomotor activity was recorded by non-invasive measurement in a home cage using infrared sensor frames (TSE Systems). All values are mean ± s.e.m. Data sets were analysed for statistical significance using a two-tailed unpaired Student's *t*-test. All *P*-values below 0.05 were considered significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 November 2008; accepted 2 February 2009.

Published online 22 February 2009.

1. Frayling, T. M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
2. Dina, C. *et al.* Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nature Genet.* **39**, 724–726 (2007).
3. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
4. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
5. Peters, T., Ausmeier, K. & Rüther, U. Cloning of *Fatso* (*Fto*), a novel gene deleted by the Fused toes (*Ft*) mouse mutation. *Mamm. Genome* **10**, 983–986 (1999).
6. Peters, T., Ausmeier, K., Dildrop, R. & Rüther, U. The mouse Fused toes (*Ft*) mutation is the result of a 1.6-Mb deletion including the entire Iroquois B gene cluster. *Mamm. Genome* **13**, 186–188 (2002).
7. van der Hoeven, F. *et al.* Programmed cell death is affected in the novel mouse mutant Fused toes (*Ft*). *Development* **120**, 2601–2607 (1994).
8. Gerken, T. *et al.* The obesity-associated *FTO* gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007).
9. Sanchez-Pulido, L. & Andrade-Navarro, M. A. The *FTO* (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem.* **8**, 23 (2007).
10. Stratigopoulos, G. *et al.* Regulation of *Fto/Ftm* gene expression in mice and humans. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294**, R1185–R1196 (2008).
11. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
12. Caro, J. F., Sinha, M. K., Kolaczynski, J. W., Zhang, P. L. & Considine, R. V. Leptin: the tale of an obesity gene. *Diabetes* **45**, 1455–1462 (1996).
13. Considine, R. V. *et al.* Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N. Engl. J. Med.* **334**, 292–295 (1996).
14. Waki, H. *et al.* Impaired multimerization of human adiponectin mutants associated with diabetes. Molecular structure and multimer formation of adiponectin. *J. Biol. Chem.* **278**, 40352–40363 (2003).
15. Lara-Castro, C., Luo, N., Wallace, P., Klein, R. L. & Garvey, W. T. Adiponectin multimeric complexes and the metabolic syndrome trait cluster. *Diabetes* **55**, 249–259 (2006).
16. Pajvani, U. B. *et al.* Complex distribution, not absolute amount of adiponectin, correlates with thiazolidinedione-mediated improvement in insulin sensitivity. *J. Biol. Chem.* **279**, 12152–12162 (2004).
17. Xu, A. *et al.* Testosterone selectively reduces the high molecular weight form of adiponectin by inhibiting its secretion from adipocytes. *J. Biol. Chem.* **280**, 18073–18080 (2005).

18. Coll, A. P., Farooqi, I. S. & O'Rahilly, S. The hormonal control of food intake. *Cell* **129**, 251–262 (2007).
19. Fredriksson, R. *et al.* The obesity gene, *FTO*, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology* **149**, 2062–2071 (2008).
20. Wilson, S. W. & Houart, C. Early steps in the development of the forebrain. *Dev. Cell* **6**, 167–181 (2004).
21. Jeon, J. Y. *et al.* *MCH*^{−/−} mice are resistant to aging-associated increases in body weight and insulin resistance. *Diabetes* **55**, 428–434 (2006).
22. Wardle, J. *et al.* Obesity associated genetic variation in *FTO* is associated with diminished satiety. *J. Clin. Endocrinol. Metab.* **93**, 3640–3643 (2008).
23. Speakman, J. R., Rance, K. A. & Johnstone, A. M. Polymorphisms of the *FTO* gene are associated with variation in energy intake, but not energy expenditure. *Obesity (Silver Spring)* **16**, 1961–1965 (2008).
24. Timpson, N. J. *et al.* The fat mass- and obesity-associated locus and dietary intake in children. *Am. J. Clin. Nutr.* **88**, 971–978 (2008).
25. Grunnet, L. G. *et al.* Increased recovery rates of phosphocreatine and inorganic phosphate after isometric contraction in oxidative muscle fibres and elevated hepatic insulin resistance in homozygous carriers of the A-allele of *FTO* rs9939609. *J. Clin. Endocrinol. Metab.* **94**, 596–602 (2009).
26. Lowell, B. B. & Bachman, E. S. β -Adrenergic receptors, diet-induced thermogenesis and obesity. *J. Biol. Chem.* **278**, 29385–29388 (2003).
27. Miyoshi, H. *et al.* Hormonal control of substrate cycling in humans. *J. Clin. Invest.* **81**, 1545–1555 (1988).
28. Wolfe, R. R., Herndon, D. N., Jahoor, F., Miyoshi, H. & Wolfe, M. H. Effect of severe burn injury on substrate cycling by glucose and fatty acids. *N. Engl. J. Med.* **317**, 403–408 (1987).
29. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
30. Vierkotten, J., Dildrop, R., Peters, T., Wang, B. & R  ther, U. Ftm is a novel basal body protein of cilia involved in Shh signalling. *Development* **134**, 2569–2577 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank S. Kuschel and B. Hampel for technical assistance, S. Fischer for critical reading of this manuscript, as well as W. Stoffel, B. Jehnke and H. Br  nneke for support during assessment of energy expenditure. This work was supported by the Deutsche Forschungsgemeinschaft (to U.R. and J.C.B.) and in part by NGFNplus (to U.R.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to U.R. (ruether@uni-duesseldorf.de) or to J.C.B. (jens.brueening@uni-koeln.de).

METHODS

Gene targeting and genotyping. For the targeted deletion of *Fto* we designed a construct to introduce a neomycin resistance cassette resulting in the deletion of exons 2 and 3 of the *Fto* coding sequence (SpeI restriction site (position 93924726 in Mm8) to BamHI restriction site (position 93934002 in Mm8)). The linearized vector was electroporated into R1 embryonic stem cells²⁹; positive clones were selected by G418 application and tested by Southern blot analysis. Targeted embryonic stem cells were injected into blastocysts (C3H×C57BL6) to create chimaeras that transmit the *Fto* mutation to their progeny on a C57BL6 background. F₁ animals were intercrossed to derive homozygous mutant F₂ mice. Genotyping was carried out by PCR analysis of DNA extracted from the tails or the yolk sac of embryos.

Animal care. Mice were housed in a virus-free facility at 22–24 °C on a 12-h light/12-h dark cycle, and fed standard rodent chow or a high-fat diet (Ssniff). All animals had access to water *ad libitum*. Food intake was measured daily over a period of 7 days. All animal procedures and euthanasia were reviewed by the animal care committee, approved by local government authorities and were in accordance with National Institutes of Health guidelines.

Blood analysis. Blood glucose was determined from whole blood using an automatic glucose monitor (GlucoMen GlycO; A. Menarini Diagnostics). Hormone levels in serum/plasma were measured by ELISA according to manufacturer's guidelines (mouse/rat insulin ELISA and leptin ELISA, Crystal Chem Inc.; quantikine mouse adiponectin/Acrp30 ELISA, R&D Systems; adrenaline research ELISA and noradrenaline research ELISA, DRG Instruments GmbH; human-free T3 ELISA kit, Alpha Diagnostic).

Glucose and insulin tolerance test. Glucose-tolerance tests were performed on animals fasted for 16 h, insulin-tolerance tests on random-fed mice. Animals were injected with either 2 g kg⁻¹ body weight of glucose, or 0.75 U kg⁻¹ body weight of human regular insulin (Actrapid, Novo Nordisk) into the peritoneal cavity and blood glucose levels were determined.

Insulin signalling. Mice were anesthetized by intraperitoneal injection of avertin (tribromoethyl alcohol and tert-amyl alcohol, Sigma-Aldrich). Samples (125 µl) of 0.9% saline with/without 5 U regular human insulin (Actrapid, Novo Nordisk) were injected into the inferior vena cava. Samples were collected 2 and 5 min after injection.

Protein extraction. Tissues were thawed and homogenized in lysis buffer (50 mM HEPES (pH 7.4), 1% Triton-X-100, 0.1 M sodium fluoride, 10 mM EDTA, 50 mM sodium chloride, 10 mM sodium orthovanadate, 0.1% SDS, 10 µg ml⁻¹ aprotinin, 2 mM benzamidine, 2 mM PMSF) using a polytron homogenizer (IKA Werke) followed by centrifugation for 1 h at 4 °C. Protein concentrations were determined by Bradford assay.

Western blot analysis. Proteins were separated by SDS-PAGE and blotted onto PVDF membranes (Bio-Rad). Membranes were incubated with 1% blocking reagent (Roche) and then with primary antibody diluted in 0.5% blocking solution for 1 h at room temperature or overnight at 4 °C. After two washing steps with TBS-Tween, membranes were incubated twice for 10 min with 0.5% blocking solution and for 1 h at room temperature with the respective secondary antibodies. The signals were visualized after four washing steps with TBS-Tween by Pierce ECL Western Blotting Substrate (Perbio Science) and exposition to chemiluminescent

film (Amersham). Bands were measured in intensity per mm² using the Quantity One Software (Bio-Rad).

mRNA expression. Total RNA for each tissue was quantified by spectrophotometry after purification using the Qiagen RNeasy Kit and DNase digest (Promega GmbH), reversely transcribed (Eurogentec), then PCR-amplified using TaqMan Principles ABI Prism 7700 Sequence Detection System with probes from Applied Biosystems. Relative expression of analysed mRNAs was determined using standard curves based on cDNA derived from the respective tissues. Samples were adjusted for total RNA content by glucuronidase, *Hprt1* and *Tfrc* RNA quantitative PCR. Calculations were performed by a comparative method ($2^{-\Delta\Delta C_t}$). *Fto* and *Ftm* expression levels were tested by semiquantitative PCR using the following primers: *Ftm*, CTCAGATCAGCCCAGCAGAGT, ACTTGGTGGAGGTAGGTGAAT; *Fto*, AGCAGAGCAGCCTACAACGTG, CTAGGATCTTGCTTCCAGCAG; and *Hprt1*, GCTGGTGAAAAGGACCTCT, CACAGGACTAGAACACCTGC.

Immunofluorescence. Cryosections were washed with PBS and permeabilized with PBS with 0.5% Triton-X-100. Blocking was performed with 10% FCS in PBS with 0.1% Triton-X-100 (FCS/PBT). The sections were incubated with the antibodies diluted in FCS/PBT overnight (primary) and for 2 h (secondary) with three washing steps after the incubation with the primary antibodies. They were then washed again and embedded in Mowiol containing DAPI (Merck).

Antibodies. The polyclonal antibodies against *Fto* were raised in rabbits immunized with recombinant *Fto* (Pineda antibody services). The antibodies were affinity-purified with the antigen coupled to *N*-hydroxysuccinimide-activated sepharose 4 fast flow (GE Healthcare). Primary antibodies for the western blot analysis were against IR- β -subunit (Santa Cruz Biotechnology Inc.), phosphorylated IR (Stressgen), AKT (Cell Signaling Technology Inc.). Mouse antibody against acetylated α -tubulin or β -actin were from Sigma-Aldrich, and secondary antibodies conjugated to HRP and to Cy2 and Cy3 were from DAKO and Dianova, respectively.

Tissue processing and histochemistry. After dissection in PBS, tissues were fixed in 4% paraformaldehyde overnight and washed in PBS. They were incubated in 70, 80, 90 and 100% ethanol for 2 h each, in 1-butanol overnight and then transferred to paraffin for embedding. Staining on paraffin sections was performed with haematoxylin and eosin or Cresylviolet (Nissl staining). Whole-mount RNA *in situ* hybridization was performed as described previously³⁰.

Analysis of body composition. Whole body composition of live animals was determined using the NMR Analyser Minispec (Bruker Optik).

Indirect calorimetry. Indirect calorimetry was measured in a Calorimetry Module (CaloSys V2.1, TSE Systems). After 2 h of acclimatization, parameters of indirect calorimetry were measured for 24 h.

Physical activity. Locomotor activity was recorded by non-invasive measurement in a home cage using infrared sensor frames (TSE Systems). The sensor frames are equipped with 16 infrared sensors arranged along the bottom of the longitudinal sides of the cage. The distance between light barriers consists of 15 mm. Locomotor activity was monitored automatically for a time period of 60 h and was quantified as the interruption of two opposed infrared diodes. Interruptions of infrared sensor pairs were detected by a control unit and registered by a computer with the relevant software (ActiMot2, TSE Systems).

Statistical methods. Unless stated otherwise, all values are mean and s.e.m. Data sets were analysed for statistical significance using a two-tailed unpaired Student's *t*-test. All *P*-values below 0.05 were considered significant.

Identification of a dendritic cell receptor that couples sensing of necrosis to immunity

David Sancho^{1*}, Olivier P. Joffre^{1*}, Anna M. Keller¹, Neil C. Rogers¹, Dolores Martínez², Patricia Hernanz-Falcón¹, Ian Rosewell³ & Caetano Reis e Sousa¹

Injury or impaired clearance of apoptotic cells leads to the pathological accumulation of necrotic corpses, which induce an inflammatory response that initiates tissue repair¹. In addition, antigens present in necrotic cells can sometimes provoke a specific immune response^{2–4} and it has been argued that necrosis could explain adaptive immunity in seemingly infection-free situations, such as after allograft transplantation or in spontaneous and therapy-induced tumour rejection^{5,6}. In the mouse, the CD8 α^+ subset of dendritic cells phagocytoses dead cell remnants and cross-primed CD8 $^+$ T cells against cell-associated antigens⁷. Here we show that CD8 α^+ dendritic cells use CLEC9A (also known as DNGR-1), a recently-characterized C-type lectin^{8–10}, to recognize a preformed signal that is exposed on necrotic cells. Loss or blockade of CLEC9A does not impair the uptake of necrotic cell material by CD8 α^+ dendritic cells, but specifically reduces cross-presentation of dead-cell-associated antigens *in vitro* and decreases the immunogenicity of necrotic cells *in vivo*. The function of CLEC9A requires a key tyrosine residue in its intracellular tail that allows the recruitment and activation of the tyrosine kinase SYK, which is also essential for cross-presentation of dead-cell-associated antigens. Thus, CLEC9A functions as a SYK-coupled C-type lectin receptor to mediate sensing of necrosis by the principal dendritic-cell subset involved in regulating cross-priming to cell-associated antigens.

We and others have recently found that the *Clec9a* gene encoding the CLEC9A protein is selectively expressed at high levels by the CD8 α^+ subset of dendritic cells and its putative human equivalent^{8–10}. To identify ligand(s) for CLEC9A, we constructed a chimera comprising the extracellular domain of CLEC9A fused to CD3 ζ (also known as CD247)⁸ and expressed it in the T cell lines BWZ and B3Z¹¹. Ligand binding to the chimaeric receptor triggers signalling via ZAP70, leading to activation of an NFAT reporter in those cells¹¹. As a control, we first checked that bivalent antibodies (but not monovalent Fab fragments, see later) to CLEC9A can trigger the fusion protein CLEC9A–CD3 ζ , in soluble form or after immobilization on plastic surfaces (Supplementary Fig. 1 and ref. 8). Notably, we observed basal activation of the NFAT reporter in the reporter cell lines expressing mouse or human CLEC9A–CD3 ζ (but not expressing a control CD3 ζ chimera made with the related C-type lectin DECTIN-1, also known as CLEC7A), which correlated with the number of dead cells in the culture (Fig. 1a). The addition of ultraviolet-irradiated mouse embryonic fibroblasts (MEFs) to the reporter cells induced signalling by mouse or human CLEC9A–CD3 ζ , but not by DECTIN-1–CD3 ζ , and this was blocked by Fab fragments of anti-mouse or anti-human CLEC9A, respectively, but not by a control Fab fragment (Fig. 1b). Although cells entered apoptosis shortly after ultraviolet treatment, they did not trigger CLEC9A–CD3 ζ signalling until 4–7.5 h after

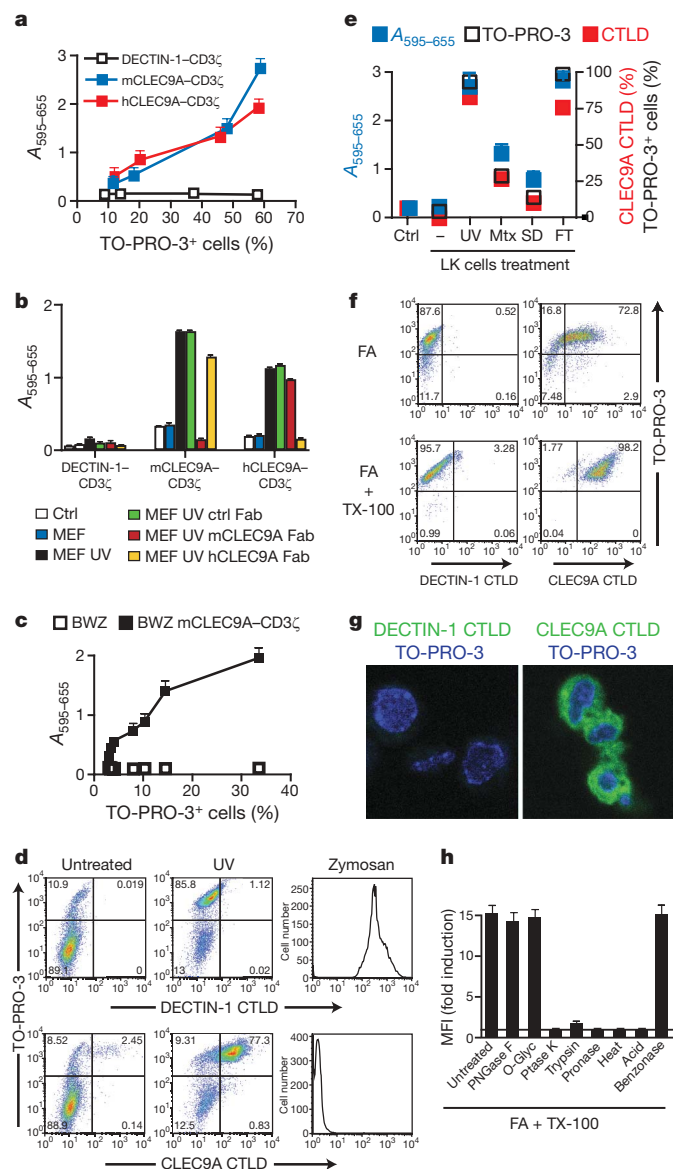
irradiation, and the signal correlated with the frequency of cells that had undergone secondary necrosis during that time period, as assessed by permeability to DNA stains such as propidium iodide and TO-PRO-3 (Fig. 1c). All cell types tested, independently of the species of origin, induced mouse and human CLEC9A–CD3 ζ signalling after ultraviolet-induced secondary necrosis (data not shown), indicating that the ligand is neither species nor cell-type restricted. As an independent means of measuring ligand expression, we made a recombinant version of the C-type lectin domain (CTLD) of CLEC9A, which contains the putative ligand binding site. Tetramers of CLEC9A CTLD, but not control tetramers of DECTIN-1 CTLD, stained the secondary necrotic (TO-PRO-3 $^+$) fraction of MEFs that had been ultraviolet-irradiated (Fig. 1d). In contrast, DECTIN-1 tetramers, but not CLEC9A tetramers, stained zymosan particles (Fig. 1d), consistent with the role of DECTIN-1 as a β -glucan-specific receptor¹². CLEC9A ligand exposure was seen after treatment with further primary or secondary necrosis-inducing protocols—including freeze-thawing, treatment with anthracyclins (mitoxantrone) or serum deprivation—and was directly proportional to the extent of TO-PRO-3 permeability induced by the treatment (Fig. 1e). Notably, cell fixation with formaldehyde, especially when followed by detergent permeabilization, rendered the cells permeable to TO-PRO-3 and instantly revealed the ligand(s) for CLEC9A (Fig. 1f–h). Further analysis showed that the ligand(s) is predominantly cytoplasmic (Fig. 1g) and resistant to glycosidase and nuclease treatment but susceptible to the action of proteases, heat and low pH (Fig. 1h). We conclude that CLEC9A recognizes a ubiquitous preformed acid-labile protein-associated ligand(s) that is normally sequestered in healthy cells but becomes exposed during necrosis, after loss of membrane integrity.

To test for a role of CLEC9A in mediating responses to necrotic cells, we generated *Clec9a^{gfp/gfp}* mice in which CLEC9A expression is abrogated by the insertion of a gene encoding a membrane-anchored form of green fluorescent protein (GFP) into the *Clec9a* open reading frame (Supplementary Fig. 2a). CD8 α^+ dendritic cells from these mice, as well as the CD8 α^+ dendritic-cell-like (CD24 hi CD11b lo B220 neg)¹³ subset from FLT3L-derived bone marrow cell cultures (FLT3L-BMDC), did not stain for CLEC9A but did express GFP (Supplementary Figs 2b, c and 3a). All other leukocytes were GFP-negative (Supplementary Fig. 2b and data not shown), consistent with the restricted expression of high levels of the *Clec9a* gene to CD8 α^+ dendritic cells^{8,10}. Dendritic cell phenotype and numbers were unaltered in *Clec9a^{gfp/gfp}* mice indicating that CLEC9A is not essential for dendritic cell development (Supplementary Fig. 2b–d), and the response of dendritic cells to innate stimuli was not altered by the loss of CLEC9A (Supplementary Figs 2f and 3b). No alterations were found in leukocyte numbers or frequencies in spleen, lymph

¹Immunobiology Laboratory, ²FACS Laboratory, Cancer Research UK, London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

³Trangen Services, Cancer Research UK, Clare Hall Laboratories, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK.

*These authors contributed equally to this work.



node or thymus (Supplementary Fig. 2b–e and data not shown), and the *Clec9a*^{gfp/gfp} mice appeared normal and did not develop any obvious symptoms of autoimmune disease (data not shown).

We first tested whether CLEC9A is necessary for the engulfment of dead cells. Using a conventional flow cytometry assay, CLEC9A⁺ and CLEC9A-deficient CD8 α ⁺-like dendritic cells from FLT3L-BMDC cultures were seen to acquire equivalent amounts of material from dye-labelled secondary necrotic splenocytes over a similar time period (Fig. 2a and Supplementary Fig. 4). We confirmed that CLEC9A-deficient CD8 α ⁺-like dendritic cells are not impaired in their ability to phagocytose necrotic cell material *in vitro* using multispectral imaging flow cytometry, which differentiates between surface binding and particle internalization (Fig. 2b). Notably, CLEC9A deficiency also did not impair uptake of cellular material by spleen CD8 α ⁺ dendritic cells *in vivo* after the injection of necrotic cells into mice (Fig. 2c). We therefore assessed whether CLEC9A might be necessary for steps subsequent to necrotic cell ingestion, culminating in the cross-presentation of dead-cell-associated antigens to CD8⁺ T cells. Ultraviolet-treated congenic (H-2^{bm1}) transformed MEFs untransfected (bm1 T) or transfected to express a non-secreted ovalbumin (OVA)-GFP fusion protein (bm1 T OVA) were allowed to undergo secondary necrosis and were then added to control (CLEC9A⁺) or *Clec9a*^{gfp/gfp} CD8 α ⁺-like dendritic cells purified from FLT3L-BMDC cultures, together with OVA-specific OT-I T-cell receptor transgenic

Figure 1 | Ligand(s) for CLEC9A are exposed after cell death. **a**, BWZ cells stably expressing mouse or human CLEC9A–CD3 ζ or DECTIN-1–CD3 ζ were plated at different concentrations and cultured for 48 h. TO-PRO-3⁺ cell frequency was determined and equivalent numbers of live cells were transferred to fresh medium. After overnight culture, LacZ activity was measured using a colourimetric assay and expressed relative to the TO-PRO-3⁺ cell frequency. **b**, Untreated or UVC-treated MEFs were cultured with BWZ cells expressing CLEC9A–CD3 ζ or DECTIN-1–CD3 ζ . Where indicated, Fab fragments of control (ctrl), anti-mCLEC9A or anti-hCLEC9A antibodies were added. Reporter activity in BWZ cells was measured as in **a**. Ctrl, BWZ cells alone. **c**, LK.35.3 (LK) cells were exposed to increasing UVC doses before culturing with BWZ cells or BWZ cells expressing mCLEC9A–CD3 ζ . LacZ activity was measured as in **a** and expressed relative to the TO-PRO-3⁺ LK cell frequency at each radiation dose. **d**, Phycoerythrin (PE)-labelled tetramers of CLEC9A or DECTIN-1 CTLD were used to stain untreated or UVC-treated MEFs or zymosan particles. **e**, BWZ cells expressing mCLEC9A–CD3 ζ were cultured overnight in medium alone (ctrl) or together with untreated (–) LK cells treated 16 h before with UVC (UV) or mitoxantrone (mtx), or serum deprived (SD) overnight, or subjected to freeze-thawing (FT). LacZ activity (left y-axis) is depicted, together with the frequency of TO-PRO-3⁺ and CLEC9A CTLD tetramer⁺ LK cells at the start of the co-culture (right y-axis). **f**, LK cells were fixed with formaldehyde (FA), and some were permeabilized with Triton X-100 (TX-100) before labelling with TO-PRO-3 and PE-tetramers of DECTIN-1 or CLEC9A CTLD. **g**, Fixed and permeabilized MEFs were labelled with DECTIN-1 or CLEC9A CTLD monomers and counterstained with TO-PRO-3 before confocal microscopy. Original magnification, $\times 630$. **h**, Fixed and permeabilized MEFs were left untreated or treated with acid (pH 3.5), peptide N-glycosidase F (PNGase F), O-glycosidase (O-glyc), proteinase K (ptase K), trypsin, pronase-E or benzoylase, or heated for 5 min at 80 °C. Cells were then labelled with DECTIN-1 or CLEC9A CTLD tetramers and analysed by flow cytometry. Data are the ratio between the mean fluorescence intensity (MFI) of the CLEC9A CTLD staining and the DECTIN-1 CTLD staining for each treatment. All data are from one representative experiment of at least three, and error bars denote the mean and s.d. of triplicate measurements.

CD8⁺ T cells. Notably, CLEC9A-deficient CD8 α ⁺-like dendritic cells showed a reduced ability to support OT-I expansion and differentiation into IFN- γ -producing effector cells, a readout for the cross-presentation of OVA (Fig. 2d, e). In contrast, when the antigen was offered in the form of a soluble protein or as a conjugate to latex beads, CLEC9A-deficient CD8 α ⁺-like dendritic cells were comparable stimulators to CLEC9A⁺ dendritic cells, indicating that CLEC9A deficiency does not intrinsically impair cross-presentation ability (Fig. 2d, e). Therefore, CLEC9A is not required for the uptake of necrotic cell material but is necessary for efficient cross-presentation of dead-cell-associated antigens by CD8 α ⁺ dendritic cells.

The CLEC9A cytoplasmic tail contains a hemITAM motif⁴ with a tyrosine at position 7, which allows binding of SYK kinase after phosphorylation (Supplementary Fig. 5a)⁹. Consistent with those data, monoclonal antibody-mediated crosslinking of CLEC9A induced phosphorylation of endogenous SYK in a CLEC9A-transfected B cell line (Supplementary Fig. 5b) and promoted activation of NFAT in B3Z cells, which was dependent on the co-expression of SYK and the presence of Tyr 7 in CLEC9A (Supplementary Figs 5c and 6). Similarly, necrotic cells triggered CLEC9A signalling in a SYK- and Tyr 7-dependent fashion (Supplementary Fig. 7a, b) independently of phagocytosis (data not shown). To address the role of signalling in dendritic cells, we first assessed whether SYK is activated in response to necrotic cells. CD8 α ⁺-like dendritic cells incubated with dead cells showed phosphorylated SYK at the interface between the dendritic cells and the corpse, which was significantly decreased when using CLEC9A-deficient cells (Fig. 3a). We then reconstituted dendritic cells from *Clec9a*^{gfp/gfp} mice with wild-type CLEC9A or the Y7F signalling-deficient mutant by retroviral transduction. Ectopic expression of the wild-type receptor restored the ability of the CD8 α ⁺-like dendritic cells to efficiently cross-present dead-cell-associated OVA, confirming that the deficiency observed in the CLEC9A-deficient cells is not due to a secondary defect unrelated to CLEC9A ablation (Fig. 3b, c). In

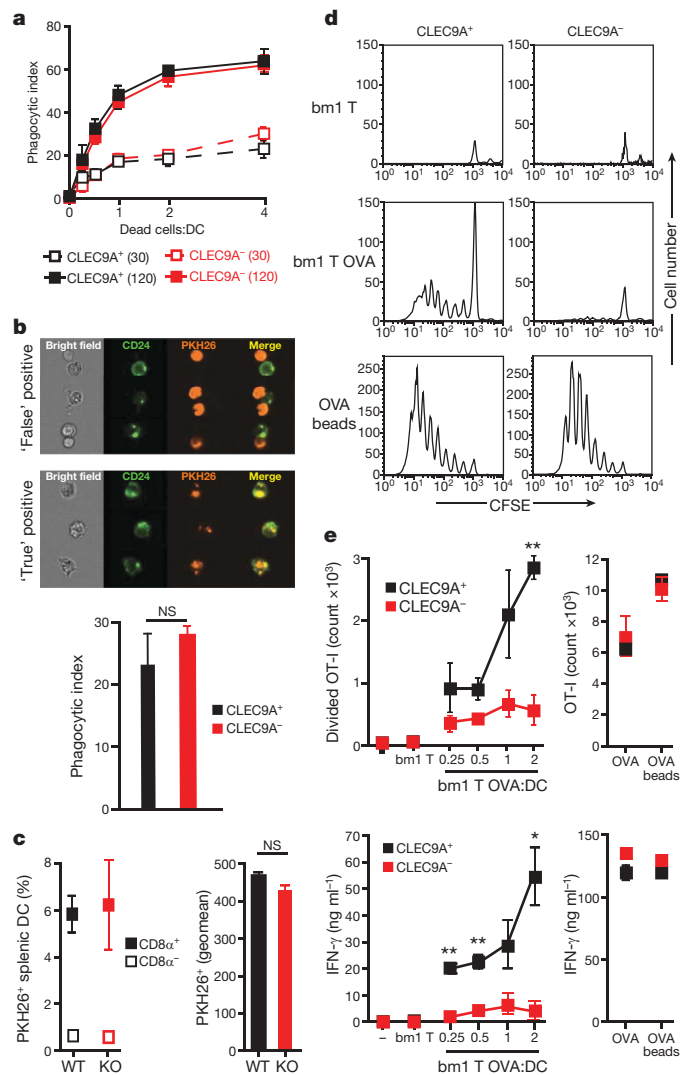


Figure 2 | CLEC9A is required for cross-presentation of dead-cell-associated antigens *in vitro*. **a**, FLT3L-BMDC from *Clec9a*^{gfp/gfp} (CLEC9A⁻) or control (CLEC9A⁺) mice were incubated at 4 °C or 37 °C for 0, 30 or 120 min with PKH26-labelled and UVC-treated H-2^{bm1} splenocytes at different ratios. The acquisition of PKH26 label by the CD8α⁺-like dendritic cell subset was quantified by flow cytometry (**a**) and normalized by subtracting the signal obtained at 4 °C from that at 37 °C. Data are mean ± s.d. of two biological replicates. The uptake of dead cell material by CD8α⁺-like dendritic cells was determined using multispectral imaging flow cytometry (**b**) to discriminate between 'true' (uptake) and 'false' (binding) positive events (**b**, top panels). The results (**b**, bottom panel) are mean and s.d. of two independent biological replicates from one experiment and are representative of two independent experiments. Original magnification in **b** was ×40. **c**, *Clec9a*^{gfp/gfp} (KO) or control *Clec9a*^{+/+} or *Clec9a*^{gfp/gfp} (WT) mice were injected with PKH26-labelled UVC-treated H-2^{bm1} splenocytes. After 2 h, the frequency (left) and the mean fluorescence (right) of CD8α⁻ and CD8α⁺ splenic dendritic cells that acquired dead cell material were determined by flow cytometry. Results are the mean and s.d. of three mice. Differences between WT and KO in **a–c** are not statistically significant. **d**, **e**, Purified CD8α⁺-like FLT3L-BMDC from *Clec9a*^{gfp/gfp} (CLEC9A⁻) or CLEC9A⁺ littermates were cultured with CFSE-labelled OVA-specific OT-I cells, and with UVC-treated H-2^{bm1} MEFs expressing (bm1 T OVA) or not expressing (bm1 T) OVA. OT-I proliferation (**d** and **e**, top panel) and IFN-γ in the supernatant (**e**, bottom panel) were quantified. Soluble OVA and OVA-coated beads were used as a control source of antigen. Results are mean ± s.e.m. of three independent biological replicates from one experiment and are representative of three independent experiments. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; Student's *t*-test.

contrast, cells transduced with the signalling-deficient receptor behaved as CLEC9A-negative control cells transduced with a virus lacking the insert and showed a profound deficiency in the cross-presentation of dead-cell-associated OVA (Fig. 3b, c). In all cases, dendritic cells cross-presented bead-associated OVA with equal efficiency (Fig. 3b, c), again indicating the selectivity of CLEC9A for dead-cell-associated antigens. We therefore assessed the ability of SYK-deficient dendritic cells to cross-present such antigens. As for CLEC9A-deficiency, loss of SYK did not impair the uptake of dead cells by CD8α⁺-like dendritic cells (Supplementary Fig. 8a) but blocked cross-presentation of dead-cell-associated OVA, but not soluble OVA, to OT-I T cells (Supplementary Fig. 8b and data not shown)¹⁵. Thus CLEC9A signalling via SYK kinase is required for cross-presentation of dead-cell-associated antigens by CD8α⁺ dendritic cells.

Furthermore, to assess the importance of CLEC9A in the cross-presentation of necrotic cell-associated antigens *in vivo*, we used a cross-priming model that relies on the intrinsic immunogenicity of dead cells to induce CTL responses to associated antigens^{2–4,16–18}. We immunized *Clec9a*^{gfp/gfp} mice and littermate controls with ultraviolet-treated bm1 T OVA cells undergoing secondary necrosis and measured anti-OVA responses 6 days later. In CLEC9A⁺ mice, these dead cells induced robust CD8⁺ T cell cross-priming as measured by specific tetramer staining or by the production of IFN-γ in response to *ex vivo* spleen cell restimulation with OVA peptide (Fig. 4a–c). Notably, cross-priming was significantly diminished in *Clec9a*^{gfp/gfp} mice when compared to littermate controls across all litters tested (Fig. 4a–c). To exclude the possibility that CLEC9A-deficient mice have an altered T cell repertoire, we sought to confirm these results in wild-type animals. We found that soluble anti-CLEC9A antibodies block the ability of dead cells to trigger mouse CLEC9A signalling (Supplementary Fig. 7), and we therefore treated C57BL/6 mice with anti-CLEC9A monoclonal antibody and measured cross-priming to dead cells expressing OVA. Anti-CLEC9A but not an isotype-matched control antibody reduced cross-priming to bm1 T OVA (Fig. 4d, e and Supplementary Fig. 9). The effect of antibody blockade was consistently more marked than that of genetic ablation (Fig. 4 and data not shown), perhaps because of redundancy in CLEC9A-deficient mice, incomplete backcrossing to the C57BL/6 strain and/or because the antibody promotes clearance of a complex of CLEC9A and other receptors involved in cross-priming to dead-cell-associated antigens. Thus, CLEC9A is necessary *in vivo* for efficient cross-priming to antigens associated with dead cells.

Here we describe CLEC9A as a CD8α⁺ dendritic-cell-expressed receptor for necrotic cells that regulates the cross-presentation of dead-cell-associated antigens by means of signalling via SYK kinase. Despite the fact that CLEC9A is an endocytic receptor^{8,9}, it does not seem to function in particle phagocytosis⁹ and, like SYK, is dispensable for the uptake of necrotic cell fragments. Rather, CLEC9A and SYK have a non-redundant role at a subsequent step(s) required for cross-presentation of dead-cell-associated antigens. We find that antibodies against CLEC9A are not delivered to lysosomal compartments upon endocytosis, unlike antibodies to CD205 (also known as LY75) (Supplementary Fig. 10a)¹⁹, and that CLEC9A co-localizes exclusively with the non-lysosomal fraction of necrotic cell material internalized by CD8α⁺ dendritic cells (Supplementary Fig. 10b). Thus, we favour a model in which CLEC9A, like the mannose receptor²⁰, diverts cargo away from lysosomal compartments to allow retrieval of antigens for cross-presentation. Because CLEC9A seems to be an activating receptor^{9,10}, similar to Dectin-1 (ref. 21), the CLEC9A-SYK pathway may also act to activate dendritic cells in response to dead cells, either autonomously²² or in collaboration with Toll-like receptors^{18,23,24}. Notably, as predicted by the 'danger' model³, CLEC9A is involved in translating innate recognition of dead cells into adaptive immunity after recognition of a preformed cellular ligand(s) that becomes exposed upon necrosis. Nevertheless, we cannot exclude the possibility that CLEC9A is also involved in cross-tolerance to dead-cell-associated antigens. Whichever the case,

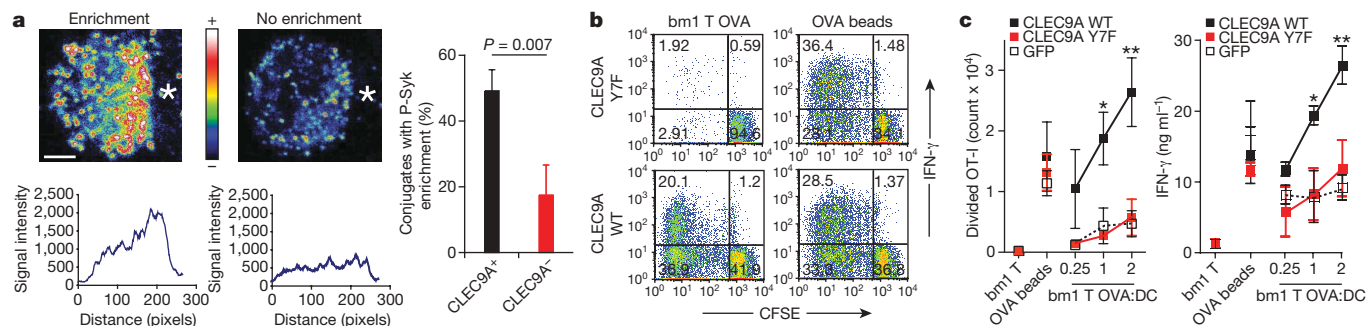


Figure 3 | CLEC9A signalling is necessary for efficient cross-presentation of dead-cell-associated antigens. **a**, CLEC9A-dependent enrichment for phospho-SYK (p-SYK) at the contact area between dendritic cells and dead cells (asterisk). Quantification of conjugates showing phospho-SYK enrichment at the contact area between CD8 α^+ -like dendritic cells and dead cells is shown as the mean and s.e.m. of three independent experiments. Scale bar, 2 μ m. **b**, **c**, CLEC9A signalling is required for efficient cross-presentation of dead-cell-associated material to CD8 $^+$ T cells. Bone marrow cells from CLEC9A $^-$ mice were retrovirally transduced with a vector encoding for GFP,

or additionally encoding for the wild-type (WT) or a mutant form of CLEC9A (Y7F). After culture in the presence of FLT3L, GFP $^+$ CD8 α^+ -like FLT3L-BMDC were purified by cell sorting and co-cultured with UVC-treated H-2 bm1 MEFs expressing (bm1 T OVA) or not expressing (bm1 T) OVA in the presence of CFSE-labelled OVA-specific OT-I cells. OT-I proliferation and IFN- γ production were measured. OVA-coated beads were used as a control source of antigen. One representative experiment out of three is shown in **b**. Data in **c** are mean \pm s.e.m. of three independent experiments. * $P < 0.05$, ** $P < 0.01$; two-way analysis of variance.

downstream of CLEC9A, our data also implicate the SRC family kinase and the SYK axis as mediators of dead cell recognition by the innate immune system. Notably, orthologues of SRC and SYK mediate clearance of apoptotic and necrotic cells by *Drosophila* glial cells²⁵, and another SYK-coupled C-type lectin was recently shown to promote inflammatory responses to necrotic cells²⁶. Thus, CLEC9A may be one of a family of SYK-coupled receptors that tap into an evolutionarily ancient mechanism for recognizing cell death.

METHODS SUMMARY

Mice. The generation of the *Clec9a* knockout mice is described in the Supplementary Methods. *Clec9a* $^+$ and *Clec9a* $^{gfp/gfp}$ mice were on a mixed 129/Sv \times C57BL/6 genetic background at the third generation of backcrossing to C57BL/6. All experiments and analyses were carried out using littermate comparisons. **Cross-presentation in vitro and in vivo.** H-2 bm1 MEFs were immortalized with SV40 T large antigen (bm1 T MEFs) and transduced to express a truncated and non-secreted OVA-GFP fusion protein. Before the assay, MEFs were ultraviolet C (UVC)-irradiated and cultured overnight in complete medium to induce secondary necrosis, and are referred to as UVC-treated cells in the Figure legends. For cross-presentation assays *in vitro*, UVC-treated bm1 T OVA MEFs were cultured with purified CD8 α^+ -like FLT3L-BMDC and carboxyfluorescein succinimidyl ester (CFSE)-labelled OVA-specific OT-I T cells. OT-I responses were quantified 4 days later by analysing IFN- γ staining and CFSE dilution profiles by flow cytometry and by monitoring IFN- γ in supernatants. For *in vivo* experiments, UVC-treated bm1 T OVA MEFs were injected intravenously (i.v.) into *Clec9a* $^{gfp/gfp}$ or control *CLEC9A* $^+$ littermates or, alternatively, C57BL/6 mice were pre-treated with an intraperitoneal (i.p.) injection of PBS, 400 μ g isotype control (rat IgG1) or 1F6 anti-CLEC9A. CD8 $^+$ T cell responses were measured 6 days later by quantifying the number of H-2K b -OVA tetramer positive cells and by IFN- γ production in response to OVA peptide restimulation.

Received 20 November; accepted 23 December 2008.

Published online 15 February 2009; corrected 16 April 2009 (details online).

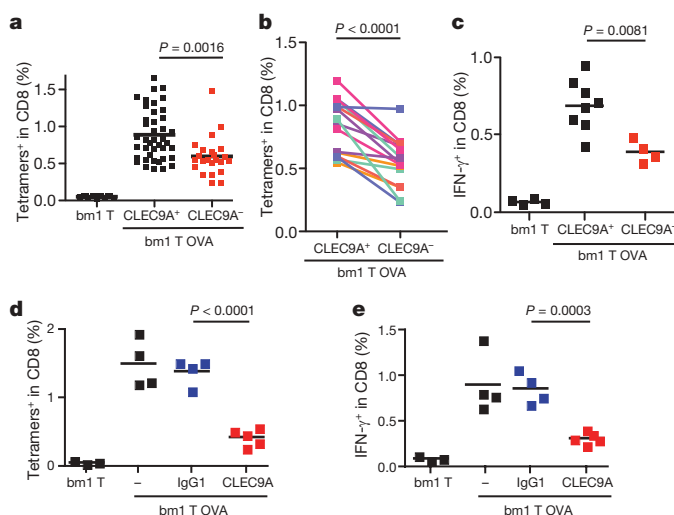


Figure 4 | CLEC9A is necessary for cross-priming to dead-cell-associated antigens in vivo. **a–c**, *Clec9a* $^{gfp/gfp}$ (CLEC9A $^-$) or control *Clec9a* $^{+/+}$ or *Clec9a* $^{gfp/+}$ (CLEC9A $^+$) littermates were immunized i.v. with 7.5×10^5 UVC-treated bm1 T OVA MEFs. Six days later, the frequency of OVA-specific CD8 $^+$ T cells was determined in the spleen (**a**, **b**). Each dot in **a** corresponds to an individual mouse; the data are pooled from seven independent experiments. The average response in each litter is shown in **b** with lines linking sex-matched (female) littermates. **c**, IFN- γ production in response to OVA peptide restimulation *ex vivo*. Data indicate the frequency of IFN- γ^+ cells within the CD8 $^+$ T cell population. Data show individual mice and the average response from three different litters. **d**, **e**, C57BL/6 mice untreated (–) or treated with anti-CLEC9A antibody or an isotype-matched control (IgG1) were immunized i.v. with 7.5×10^5 UVC-treated bm1 T OVA MEFs. The frequency of OVA-specific CD8 $^+$ T cells (**d**) or of IFN- γ producing CD8 $^+$ T cells (**e**) in response to OVA peptide restimulation was measured 6 days later. Individual mice and the average for one representative experiment out of three are shown. P values were determined using the Student's t -test (paired for **b**).

- Rock, K. L. & Kono, H. The inflammatory response to cell death. *Annu. Rev. Pathol.* 3, 99–126 (2008).
- Gallucci, S., Lolkema, M. & Matzinger, P. Natural adjuvants: endogenous activators of dendritic cells. *Nature Med.* 5, 1249–1255 (1999).
- Shi, Y., Zheng, W. & Rock, K. L. Cell injury releases endogenous adjuvants that stimulate cytotoxic T cell responses. *Proc. Natl Acad. Sci. USA* 97, 14590–14595 (2000).
- Sauter, B. *et al.* Consequences of cell death. Exposure to necrotic tumor cells, but not primary tissue cells or apoptotic cells, induces the maturation of immunostimulatory dendritic cells. *J. Exp. Med.* 191, 423–434 (2000).
- Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* 12, 991–1045 (1994).
- Matzinger, P. The danger model: a renewed sense of self. *Science* 296, 301–305 (2002).
- Villadangos, J. A. & Schnorrer, P. Intrinsic and cooperative antigen-presenting functions of dendritic-cell subsets *in vivo*. *Nature Rev. Immunol.* 7, 543–555 (2007).
- Sancho, D. *et al.* Tumor therapy in mice via antigen targeting to a novel, DC-restricted C-type lectin. *J. Clin. Invest.* 118, 2098–2110 (2008).
- Huysamen, C., Willment, J. A., Dennehy, K. M. & Brown, G. D. CLEC9A is a novel activation C-type lectin-like receptor expressed on BDCA3 $^+$ dendritic cells and a subset of monocytes. *J. Biol. Chem.* 283, 16693–16701 (2008).
- Caminschi, I. *et al.* The dendritic cell subtype restricted C-type lectin Clec9A is a target for vaccine enhancement. *Blood* 112, 3264–3273 (2008).
- Karttunen, J., Sanderson, S. & Shastri, N. Detection of rare antigen-presenting cells by the lacZ T-cell activation assay suggests an expression cloning strategy for T-cell antigens. *Proc. Natl Acad. Sci. USA* 89, 6020–6024 (1992).

12. Brown, G. D. & Gordon, S. Immune recognition. A new receptor for β -glucans. *Nature* **413**, 36–37 (2001).
13. Naik, S. H. *et al.* Cutting edge: generation of splenic CD8⁺ and CD8[−] dendritic cell equivalents in Fms-like tyrosine kinase 3 ligand bone marrow cultures. *J. Immunol.* **174**, 6592–6597 (2005).
14. Robinson, M. J., Sancho, D., Slack, E. C., LeibundGut-Landmann, S. & Reis e Sousa, C. Myeloid C-type lectins in innate immunity. *Nature Immunol.* **7**, 1258–1265 (2006).
15. LeibundGut-Landmann, S., Osorio, F., Brown, G. D. & Reis e Sousa, C. Stimulation of dendritic cells via the dectin-1/Syk pathway allows priming of cytotoxic T-cell responses. *Blood* **112**, 4971–4980 (2008).
16. Casares, N. *et al.* Caspase-dependent immunogenicity of doxorubicin-induced tumor cell death. *J. Exp. Med.* **202**, 1691–1701 (2005).
17. Obeid, M. *et al.* Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nature Med.* **13**, 54–61 (2006).
18. Apetoh, L. *et al.* Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nature Med.* **13**, 1050–1059 (2007).
19. Mahnke, K. *et al.* The dendritic cell receptor for endocytosis, DEC-205, can recycle and enhance antigen presentation via major histocompatibility complex class II-positive lysosomal compartments. *J. Cell Biol.* **151**, 673–684 (2000).
20. Burgdorf, S., Kautz, A., Böhnert, V., Knolle, P. A. & Kurts, C. Distinct pathways of antigen uptake and intracellular routing in CD4 and CD8 T cell activation. *Science* **316**, 612–616 (2007).
21. LeibundGut-Landmann, S. *et al.* Syk- and CARD9-dependent coupling of innate immunity to the induction of T helper cells that produce interleukin 17. *Nature Immunol.* **8**, 630–638 (2007).
22. Janssen, E. *et al.* Efficient T cell activation via a Toll-interleukin 1 receptor-independent pathway. *Immunity* **24**, 787–799 (2006).
23. Kim, H. S. *et al.* Toll-like receptor 2 senses β -cell death and contributes to the initiation of autoimmune diabetes. *Immunity* **27**, 321–333 (2007).
24. Cavassani, K. A. *et al.* TLR3 is an endogenous sensor of tissue necrosis during acute inflammatory events. *J. Exp. Med.* **205**, 2609–2621 (2008).
25. Ziegenfuss, J. S. *et al.* Draper-dependent glial phagocytic activity is mediated by Src and Syk family kinase signalling. *Nature* **453**, 935–939 (2008).
26. Yamasaki, S. *et al.* Mincle is an ITAM-coupled activating receptor that senses damaged cells. *Nature Immunol.* **9**, 1179–1188 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by Cancer Research UK. D.S. was supported by an EMBO long-term fellowship (ALTF 336-2004) and by a Marie Curie Intra-European Fellowship within the sixth European Community Framework Programme (MEIF-CT-2005-009205). We thank E. Schweighoffer and V. Tybulewicz for fetal liver from Syk^{−/−} embryos. We also thank N. Shastri for B3Z and BWZ cell lines and A. Eddaoudi for assistance with cell sorting. We are grateful to members of the Immunobiology Laboratory, Cancer Research UK, for advice and discussions and the Biological Resources staff for animal care and assistance with mouse experiments.

Author Contributions D.S. and O.P.J. performed most of the experiments. A.M.K. performed the phospho-SYK staining and analysis of CLEC9A subcellular distribution. D.S., O.P.J. and C.R.S. planned the research, analysed and interpreted data and wrote the manuscript. N.C.R. helped with mouse breeding and genotyping. D.M. performed and analysed cell sorting and multispectral flow cytometry experiments. P.H.-F. generated SYK transfectants. I.R. contributed to the generation of CLEC9A-deficient mice.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.R.S. (caetano@cancer.org.uk).

LETTERS

IFN α activates dormant haematopoietic stem cells *in vivo*

Marieke A. G. Essers^{1,2}, Sandra Offner³, William E. Blanco-Bose³, Zoe Waibler⁴, Ulrich Kalinke^{4,5}, Michel A. Duchosal⁶ & Andreas Trumpp^{1,2,3}

Maintenance of the blood system is dependent on dormant haematopoietic stem cells (HSCs) with long-term self-renewal capacity. After injury these cells are induced to proliferate to quickly re-establish homeostasis¹. The signalling molecules promoting the exit of HSCs out of the dormant stage remain largely unknown. Here we show that in response to treatment of mice with interferon- α (IFN α), HSCs efficiently exit G₀ and enter an active cell cycle. HSCs respond to IFN α treatment by the increased phosphorylation of STAT1 and PKB/Akt (also known as AKT1), the expression of IFN α target genes, and the upregulation of stem cell antigen-1 (Sca-1, also known as LY6A). HSCs lacking the IFN α / β receptor (IFNAR)², STAT1 (ref. 3) or Sca-1 (ref. 4) are insensitive to IFN α stimulation, demonstrating that STAT1 and Sca-1 mediate IFN α -induced HSC proliferation. Although dormant HSCs are resistant to the anti-proliferative chemotherapeutic agent 5-fluoro-uracil^{1,5}, HSCs pre-treated (primed) with IFN α and thus induced to proliferate are efficiently eliminated by 5-fluoro-uracil exposure *in vivo*. Conversely, HSCs chronically activated by IFN α are functionally compromised and are rapidly out-competed by non-activatable *Ifnar*^{-/-} cells in competitive repopulation assays. Whereas chronic activation of the IFN α pathway in HSCs impairs their function, acute IFN α treatment promotes the proliferation of dormant HSCs *in vivo*. These data may help to clarify the so far unexplained clinical effects of IFN α on leukaemic cells^{6,7}, and raise the possibility for new applications of type I interferons to target cancer stem cells⁸.

Interferon (IFN) cytokines are produced by cells of the immune system in response to challenges by agents such as viruses, bacteria and tumour cells. IFNs suppress viral replication, have immunomodulatory activities and are used clinically to treat viral diseases, multiple sclerosis and malignancies such as chronic myeloid leukaemia (CML)^{6,9,10}. Type I interferons (IFN α , IFN β) bind and signal through the IFN α / β receptor (IFNAR)¹¹ and are strongly induced by the viral genome of many RNA viruses, which can be mimicked by the double-stranded RNA mimetic polyinosinic-polycytidylic acid (hereafter termed poly(I:C))¹². Because *Mx* genes are strongly induced by IFNs, the *Mx1* promoter was used to generate the first inducible transgenic Cre line, *Mx1-Cre*, now one of the most commonly used mouse lines to eliminate genes flanked by *loxP* sites (floxed) in HSCs¹³. Although all functional mouse HSCs are contained within the Lin^{neg} Sca1⁺ c-kit⁺ (LSK) population, less than 10% of LSK cells are repopulating HSCs. Significantly higher stem cell purities can be achieved by selecting LSK cells that are CD34⁻, that show differential expression of SLAM receptors (CD150⁺ CD48⁻), or by selecting dormant HSCs using label-retaining assays^{1,14,15}.

Analysis of HSCs of 'control' animals lacking the *Mx1-Cre* or floxed genes, with or without poly(I:C) treatment, indicated that the proliferation of functional HSCs may be affected by poly(I:C) or IFN α . To systematically investigate this initial observation, C57Bl/6 (wild-type) mice were injected with poly(I:C) and the proliferation of the HSC/progenitor compartments was assessed by BrdU incorporation assays. As shown in Fig. 1a, proliferation of LSKCD150⁺ cells is strongly induced in wild-type mice, but not in mice lacking the IFNAR, showing that poly(I:C)-mediated HSC proliferation requires IFNAR signalling. As expected, poly(I:C) treatment causes a significant increase in serum IFN α (Supplementary Fig. 1a). Moreover, poly(I:C)-induced proliferation is not restricted to LSKCD150⁺ cells, but also occurs in populations further enriched in functional, predominantly quiescent HSCs such as LSKCD150⁺ CD34⁻ or LSKCD135⁻ CD34⁻ subsets (Supplementary Fig. 1b)^{14–16}. The effects of poly(I:C) on proliferation of LSKCD150⁺ cells is transient, peaking around 48 h after injection, and returning to normal after 4 days (Supplementary Fig. 1c). To examine whether IFN α has the same effect as poly(I:C), wild-type mice were injected with mouse recombinant IFN α 4. Doses above 1,000 units per mouse significantly increased the proliferation of all tested HSC enriched populations, suggesting that not only HSCs but also early progenitors are stimulated to cycle in response to IFN α (Fig. 1b, c and Supplementary Fig. 1d, e).

As expected, LSKCD150⁺ CD48⁻ HSCs in untreated mice are predominantly in a quiescent, intracellular Ki67 negative (icKi67^{neg} Hoechst^{low}) G₀ phase¹⁵. However, within 16 h of poly(I:C) injection most of these cells exit G₀ and enter an active cell cycle state (icKi67⁺ Hoechst^{med-hi}) (Fig. 1d, e)^{15,17}. To examine whether IFN α signalling also activates highly functional dormant HSCs, long-term label-retaining assays were performed¹. DNA was labelled *in vivo* with BrdU for 10 days, followed by a 14-week BrdU-free 'chase' period, thus revealing long-lived dormant 'label retaining cells' (LRCs) within the LSKCD150⁺ CD48⁻ CD34⁻ population, previously shown to contain the most highly potent HSCs¹. After 12 weeks of chase, wild-type and *Ifnar*^{-/-} mice were injected with either IFN α or poly(I:C), and the percentage of LRCs in various HSC populations was determined by fluorescence-activated cell sorting (FACS) at week 14. Notably, although the total number of cells remained unaltered (Supplementary Fig. 1f), the proportion of LRCs within the most quiescent HSC compartment (LSKCD150⁺ CD48⁻ CD34⁻)^{1,18} was greatly diminished in wild-type mice, suggesting that these cells have divided and diluted out the BrdU-label (Fig. 1f). As expected, although *Ifnar*^{-/-} mice have a slightly reduced quiescent HSC compartment (Supplementary Fig. 1d), these cells do not proliferate in response to IFN α (Fig. 1h). These data indicate that even the most dormant HSC

¹Division of Stem Cells and Cancer, Deutsches Krebsforschungszentrum (DKFZ), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany. ²Heidelberg Institute for Stem Cell Technologies and Experimental Medicine (HI-STEM), Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany. ³Ecole Polytechnique Fédérale de Lausanne (EPFL), ISREC—Swiss Institute for Experimental Cancer Research, School of Life Science, 1015 Lausanne, Switzerland. ⁴Division of Immunology, Paul Ehrlich Institute, D-63225 Langen, Germany. ⁵TWINCORE—Centre for Experimental and Clinical Infection Research Feodor-Lynen-Str. 7, 30625 Hannover, Germany. ⁶Service and Central Laboratory of Hematology, CHUV, University Hospitals of Lausanne, CH-1011 Lausanne, Switzerland.

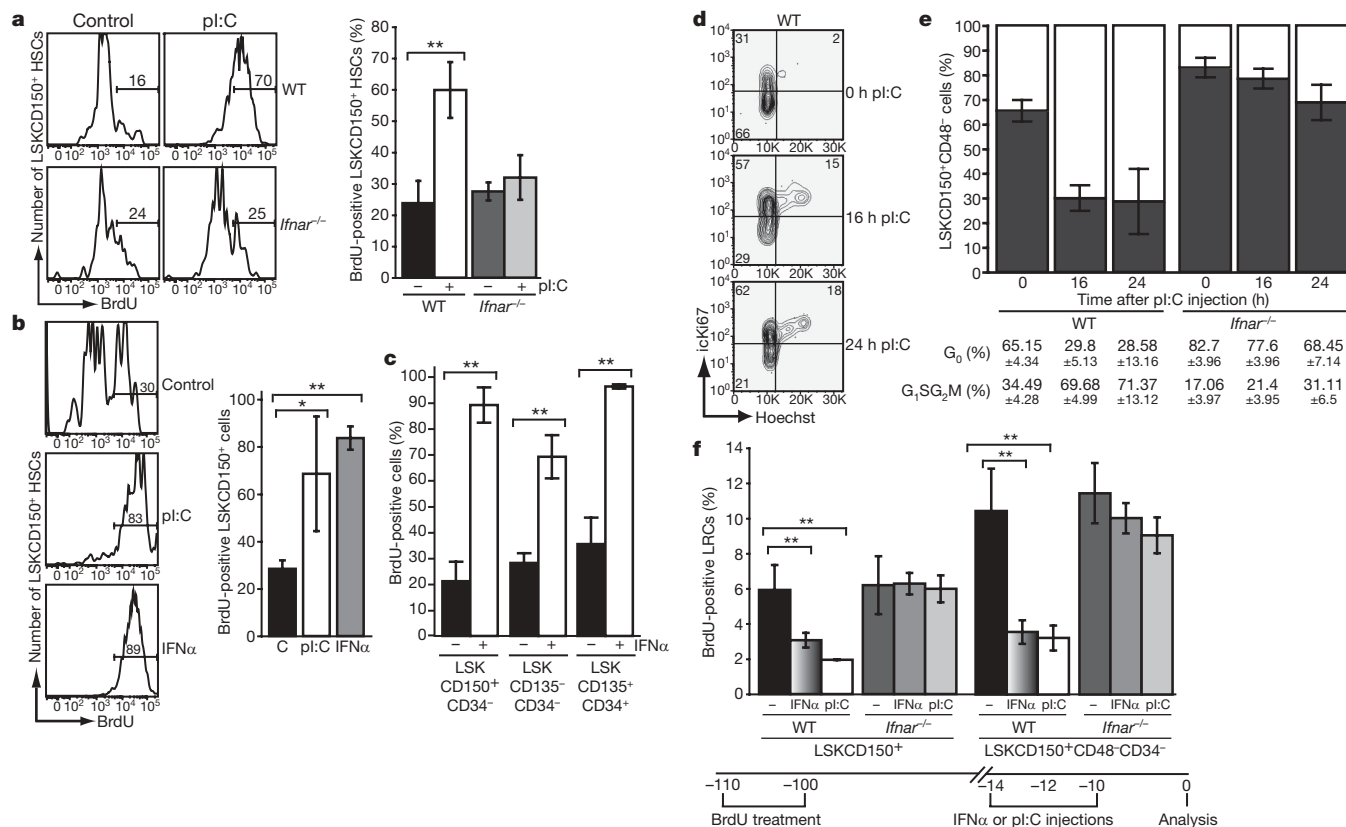


Figure 1 | IFN α induces efficient cell cycle entry of dormant HSCs *in vivo*.

a, Sixteen-hours of BrdU incorporation in LSKCD150⁺ cells from control or poly(I:C)-injected (pl:C; 10 μ g g⁻¹; 0 and 48 h) wild-type (WT) and *Ifnar*^{-/-} mice. FACS analysis at 72 h. Left, representative FACS profiles; right, quantitative and statistical analysis. **b**, Sixteen-hours of BrdU-uptake in LSKCD150⁺ cells from wild-type mice injected with poly(I:C) (10 μ g g⁻¹; 0 and 48 h), or with IFN α (10,000 units per mouse; 0, 24 and 48 h). Analysis at 72 h is shown. Left, representative FACS profiles; right, quantitative and statistical analysis. **c**, Analysis as in Supplementary Fig. 1b but using 10,000 units per mouse IFN α . **d**, Cell cycle analysis (Hoechst versus

intracellular (ic) Ki67 to detect G₀ cells, iKi67^{neg} Hoechst^{low}) on LSKCD150⁺CD48⁻ cells of wild-type mice, untreated or injected with poly(I:C) (5 μ g g⁻¹; for 16 or 24 h). **e**, Quantitative analysis of **d**. Grey bars denote percentage G₀; white bars denote percentage G₁/S/G₂/M. **f**, BrdU LRC assay on wild-type and *Ifnar*^{-/-} mice. The percentage of LRCs in each HSC-subpopulation plus poly(I:C) or mouse IFN α as indicated in the scheme. Data are the mean \pm s.d. of four mice minimum per condition. These experiments were done in triplicate with similar results. **P* < 0.05, ***P* < 0.01 (two-tailed *t*-test).

populations are induced to efficiently proliferate in response to IFN α stimulation *in vivo*.

It is thought that IFN α inhibits cellular proliferation, an observation which has been extensively demonstrated using *in vitro* culture systems¹⁰. Therefore our *in vivo* results are unexpected and raise the possibility that the observed HSC activation is caused by an IFN α -mediated effect on the stem cell niche¹⁹. However, wild-type HSCs present in chimaeric mice comprising wild-type haematopoietic cells in an *Ifnar*^{-/-} stromal environment (reverse chimaeras) are still efficiently activated by IFN α , suggesting that IFNAR signalling in the stem cell niche itself is not required (Supplementary Fig. 2a). Although IFN α stimulation is sufficient to efficiently activate the proliferation of HSCs, IFNAR-mediated signalling is not required for HSC function, as *Ifnar*^{-/-} HSCs can long-term reconstitute the entire blood system of lethally irradiated recipients (Supplementary Fig. 2b). Moreover, in mixed chimaeras generated by transplanting various input ratios of wild-type and *Ifnar*^{-/-} bone marrow (5:95%, 50:50% and 95:5%) (Fig. 2a), the contribution of the *Ifnar*^{-/-} bone marrow remains constant over a period of 3–5 months (6 \pm 2%, 49 \pm 3% and 94 \pm 3% (means \pm s.d.) *Ifnar*^{-/-} cells, respectively; *n* = 12, *P* < 0.01), suggesting that *Ifnar*^{-/-} HSCs have no competitive disadvantage over wild-type cells. To test whether IFN α directly or indirectly activates HSCs *in vivo*, 5:95% wild-type:*Ifnar*^{-/-} chimaeras were stimulated with IFN α . This resulted in the efficient proliferation of the few wild-type (CD45.1⁺) LSK-HSCs (Fig. 2b, left panel), even though 95% of the 'haematopoietic environment' is *Ifnar*^{-/-} and thus

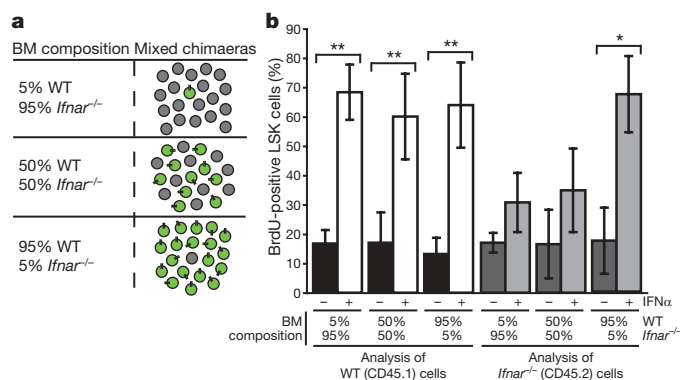


Figure 2 | A direct effect on HSCs, combined with an indirect mechanism, promotes HSC proliferation in response to IFN α *in vivo*. **a**, Schematic illustration of the generation and analysis of mixed bone marrow (BM) chimaeras. **b**, Proliferation of LSK cells in chimaeras comprised of different ratios of wild-type (WT; IFN α responsive) and *Ifnar*^{-/-} (IFN α non-responsive) bone marrow. On the left, analysis gating on wild-type (CD45.1⁺) cells is shown; on the right, analysis gating on *Ifnar*^{-/-} (CD45.2⁺) cells in the indicated chimaeras is shown. Each set represents six chimaeras, untreated or treated (24 and 16 h before analysis) with IFN α (10,000 units per mouse). BrdU was injected 16 h before analysis. Data are the mean \pm s.d. of four mice minimum per condition. These experiments were done in triplicate with similar results. **P* < 0.05, ***P* < 0.01 (two-tailed *t*-test).

unable to respond to IFN α . Furthermore, increasing the IFN α responsive wild-type haematopoietic environment from 5 to 95% did not further augment HSC proliferation (Fig. 2b, left panel) suggesting that IFN α directly promotes the proliferation of LSK-HSCs. In contrast, with high (but not low) numbers of wild-type cells present in the bone marrow, even *Ifnar*^{-/-} LSK-HSCs (CD45.2⁺) were induced to proliferate in response to IFN α treatment (95:5% wild-type:*Ifnar*^{-/-} chimaeras) (Fig. 2b, right panel). This indicates that besides directly activating wild-type HSCs, IFN α can also promote the proliferation of cells lacking IFNAR, but only if a sufficient number of wild-type bone marrow cells are present. This indirect effect of IFN α is probably due to a positive feedback loop caused by the inhibitory effects of IFN α on wild-type differentiated blood cells, similar to what is observed after treatment with 5-fluoro-uracil (5-FU)^{1,5,10}.

To confirm the direct effect of IFN α on HSCs, complementary DNA microarray analysis was performed on sorted Lin^{neg}cKit⁺CD150⁺CD48⁻ HSCs from IFN α -treated (16 h) and untreated mice. This analysis showed the specific induction of a typical set of known IFN target genes in HSCs from IFN α treated mice, strongly supporting our data suggesting that IFN α directly activates dormant HSCs (Supplementary Table 1). Furthermore, the expression of several cell-cycle genes is altered in IFN α -stimulated HSCs, including upregulation of *Ccnb2* and *Cdk7*, as well as repression of *Pten* and *Elavl1*, known to stabilize p21^{Cip1} protein²⁰.

Type I IFN signalling is known to be mediated by activation of the JAK-STAT signalling pathway^{9,11,21}. Indeed, not only are *Stat1* transcripts upregulated sixfold (Supplementary Table 1), a significant and transient increase in the phosphorylation of STAT1 (pSTAT1) is also observed in response to poly(I:C) in wild-type but not *Ifnar*^{-/-} HSCs (Fig. 3a and Supplementary Fig. 3a, b). Interestingly, HSCs from poly(I:C)-treated *Stat1*^{-/-} mice³ showed no increase in BrdU-uptake, genetically demonstrating the requirement of STAT1 signalling for IFN α -mediated HSC activation (Fig. 3b and Supplementary Fig. 3c).

In agreement with the observed IFN α -mediated repression of PTEN expression in HSCs (Supplementary Table 1), an increase of PKB/Akt phosphorylation was detected, suggesting that IFN α stimulation of HSCs results in activation of the PI3K signalling pathway (Fig. 3c). Part of this PI3K signalling seems to be downstream of STAT1, because the increase in PKB/Akt phosphorylation is reduced in *Stat1*^{-/-} mice (Figs 3d and 4g).

Unexpectedly, IFN α -treated Lin^{neg}cKit⁺CD150⁺ cells also show a significant increase in messenger RNA and cell surface expression of Sca-1 (Supplementary Table 1 and Fig. 3e). This glycosyl-phosphatidylinositol-linked cell surface receptor is highly expressed on all functional mouse HSCs, and *Sca-1*^{-/-} HSCs have a competitive disadvantage in the presence of wild-type HSCs suggesting that this protein is important for HSC self-renewal^{14,22}. To examine the role of Sca-1 in the IFN α -induced effect on HSC proliferation, poly(I:C)-treated *Sca-1*^{-/-} mice⁴ were analysed. Although the number of proliferating *Sca-1*^{-/-} HSCs during homeostasis is slightly higher compared to control mice, treatment of poly(I:C) does not further augment HSC proliferation (Fig. 3f and Supplementary Fig. 3d). Furthermore, Sca-1 is not induced after IFN α stimulation in the absence of STAT1 (Fig. 3g). These results show that Sca-1 is downstream of IFN α -STAT1 signalling, and suggests that IFN α -induced HSC proliferation is not only mediated by STAT1, but also by Sca-1 (Fig. 4e).

Proliferating, but not dormant, HSCs are sensitive to treatment with anti-proliferative agents such as 5-FU²³. Healthy mice survive even repeated treatment with 5-FU because dormant, and therefore drug-resistant HSCs are recruited into the cell cycle to rapidly produce new cells to quickly re-establish homeostasis¹. Shifting the balance of the dormant HSC pool towards self-renewal has been shown to make mice exquisitely sensitive to repeated 5-FU treatment^{24,25}. To test whether dormant, drug-resistant HSCs can be driven out of quiescence by IFN α , wild-type mice were primed at different times with poly(I:C), followed by two rounds of 5-FU

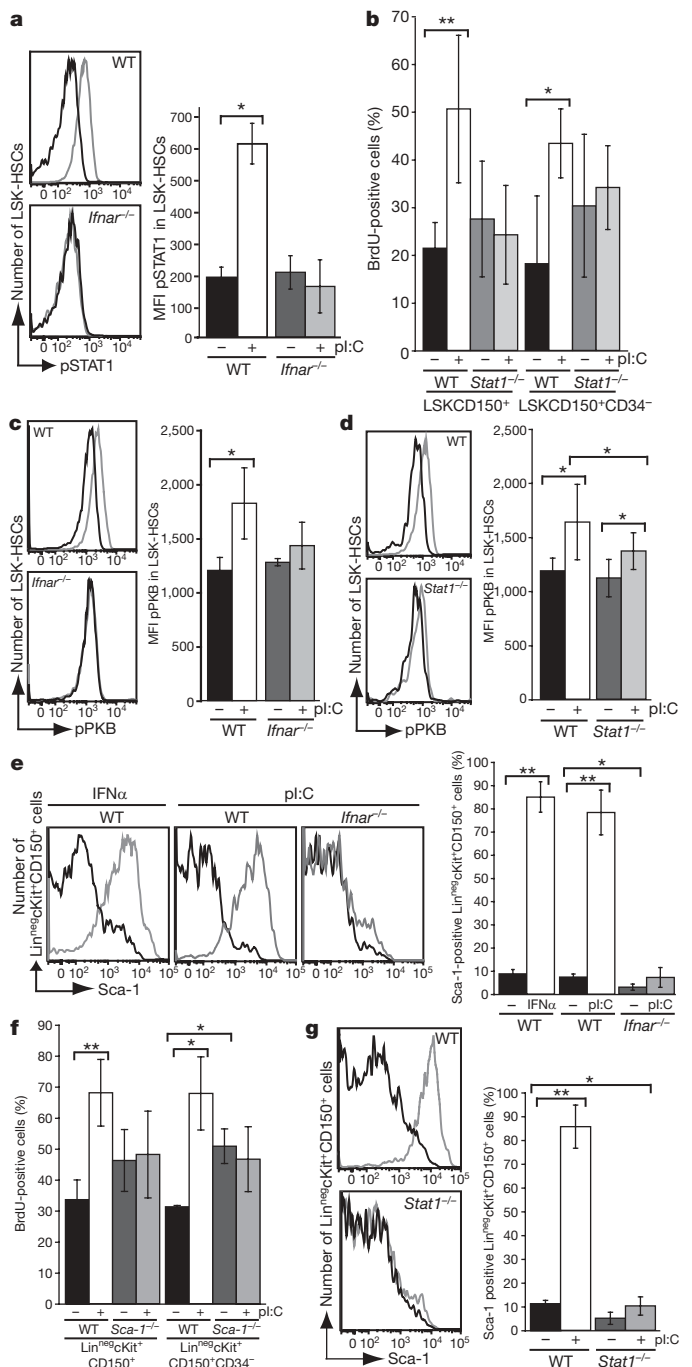


Figure 3 | IFN α -activated HSCs upregulate and are dependent on STAT1 and Sca-1 signalling. **a**, pSTAT1 expression in LSK cells of wild-type and *Ifnar*^{-/-} mice, 16 h after i.p. injection of 5 μ g g⁻¹ poly(I:C) (denoted as pl:C). Left, representative FACS profiles; right, quantitative and statistical analysis; MFI, median fluorescence intensity. **b**, The 12 h BrdU-uptake of HSCs from wild-type and *Stat1*^{-/-} mice 72 h after poly(I:C) (5 μ g g⁻¹; at 0 and 48 h). **c**, Phosphorylated PKB/Akt (pPKB) expression in LSK cells of wild-type (WT) and *Ifnar*^{-/-} mice 16 h after 5 μ g g⁻¹ poly(I:C) i.p. Left, representative FACS profiles; right, quantitative and statistical analysis. **d**, pPKB expression in LSK cells of wild-type and *Stat1*^{-/-} mice; treatment as in **c**. **e**, Sca-1 expression in Lin^{neg}cKit⁺CD150⁺ cells of wild-type and *Ifnar*^{-/-} mice after IFN α and poly(I:C) as in Fig. 1d. Left, representative FACS profiles; right, quantitative and statistical analysis. **f**, Proliferation of indicated HSC populations isolated from wild-type or *Sca-1*^{-/-} mice. Poly(I:C) injections as in **b**. **g**, The percentage of Sca-1⁺ cells within the Lin^{neg}cKit⁺CD150⁺ population in wild-type and *Stat1*^{-/-} mice, poly(I:C) injection as in **b**. Data are the mean \pm s.d. of four mice minimum per condition. **P* < 0.05, ***P* < 0.01 (two-tailed *t*-test).

treatment 7 days apart (Fig. 4a). Using this strategy, a treatment schedule was identified, resulting in the death of mice due to severe anaemia, probably because of the total loss of HSCs (Fig. 4b). Treatment schedules leading to death of the animals correlated with

the increased proliferative status of LSKCD150⁺ HSCs at the start of treatment, thus sensitizing them to 5-FU-mediated killing (Supplementary Fig. 4a). Lethality observed under these treatment regimes was dependent on IFNAR signalling because none was observed in *Ifnar*^{-/-} mice (Supplementary Fig. 4b). These data indicate that IFN α priming provides an efficient way to induce cell-cycle entry of dormant HSCs, thus making them susceptible to elimination by anti-proliferative chemotherapeutic drugs such as 5-FU.

Because extensive HSC proliferation can lead to the exhaustion of stem cell function²⁵, bone marrow isolated from mice treated three times with poly(I:C) (followed by 10 days recovery) was serially transplanted into irradiated recipients. Because no significant difference in the repopulation activity of poly(I:C) treated cells was observed, transient activation of IFN α signalling does not affect the number of functional HSCs (Supplementary Fig. 4c). To study whether long-term (chronic) activation leads to a decrease in HSC activity, the mixed chimaeras described in Fig. 2a were treated eight times with poly(I:C) (every second day) and the number of phenotypic HSCs was determined 8 days later. Notably, in 50:50 wild-type:*Ifnar*^{-/-} chimaeras all wild-type HSCs were lost and the HSCs present were exclusively derived from *Ifnar*^{-/-} donors (Fig. 4c). Moreover, chimaeras containing 95:5% wild-type:*Ifnar*^{-/-} HSCs before poly(I:C) treatment comprised only 12% wild-type and 88% *Ifnar*^{-/-} HSCs after chronic IFN α signalling suggesting a marked competitive disadvantage of IFN α -stimulated HSCs (Fig. 4d). Moreover, transplantation of bone marrow from these mice at day 22 confirmed the complete loss of functional wild-type HSCs, as recipient mice showed no bone marrow reconstitution derived from wild-type but only *Ifnar*^{-/-} cells (Supplementary Fig. 4d). Interestingly, in contrast to wild-type:*Ifnar*^{-/-} mixed chimaeras, chronic IFN α treatment of non-chimaeric wild-type mice did not show any obvious HSC phenotype (data not shown). These results indicate that long-term activation of the IFN α signalling pathway does not simply eliminate HSCs, but rather markedly compromises their function, explaining why they are rapidly out competed by *Ifnar*^{-/-} HSCs in competitive repopulation assays.

IFN α used to be a first line treatment for CML with variable outcomes. In recent years, imatinib mesylate has replaced IFN α owing to its far higher response rate accompanied with fewer side effects^{26,27}. Unfortunately, discontinuation of imatinib mesylate results in a very high relapse rate, apparently due to the imatinib-mesylate-resistance of CML stem cells^{6,28}. However, and relevant to our findings here, six patients initially treated with IFN α but subsequently switched to imatinib mesylate in 2002 showed a surprisingly high rate of long-term complete remission²⁹. The anti-CML mechanism exerted by IFN α has classically been linked to the effects on immune cells including cytotoxic T cells^{6,30}. As a non-exclusive alternative our results raise

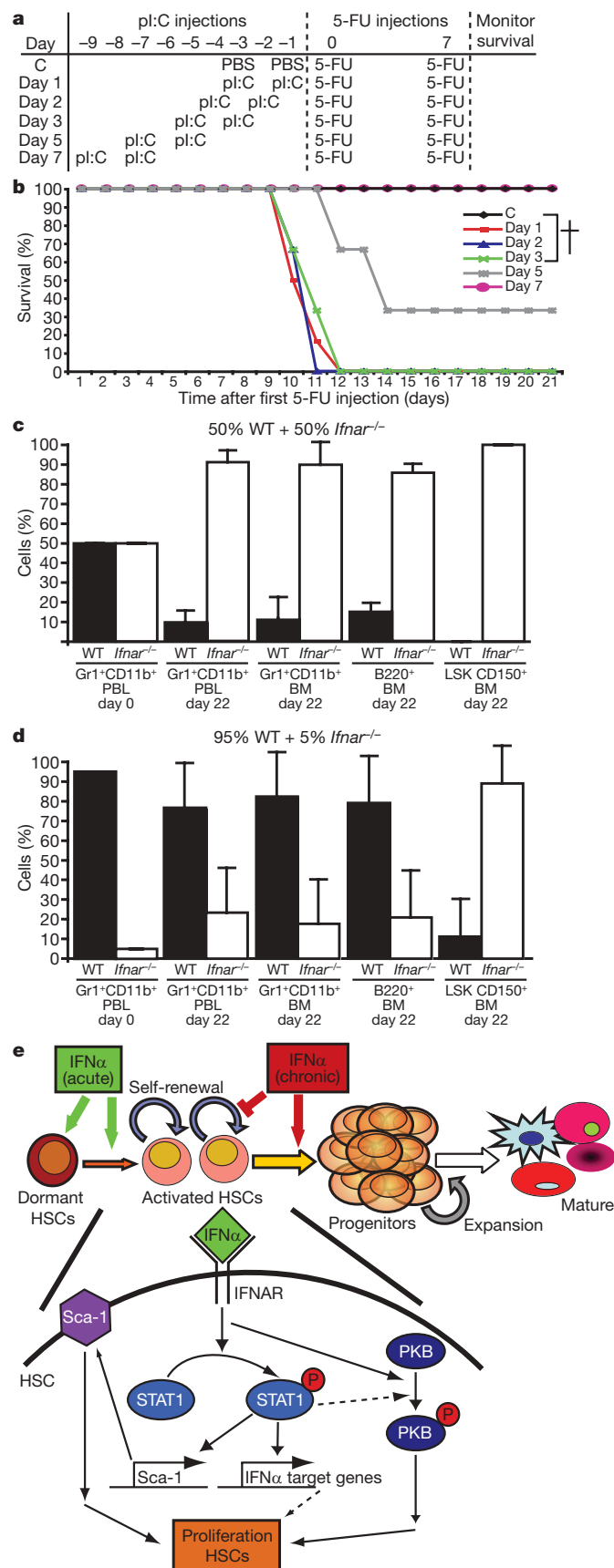


Figure 4 | Priming of IFN α followed by 5-FU treatment or chronic IFN α stimulation causes HSC loss. **a**, Schematic of various pre-treatments (priming) of wild-type (WT) mice with 10 $\mu\text{g g}^{-1}$ poly(I:C) (pI:C) or PBS before injection with 5-FU. C, control. **b**, Kaplan-Meier analysis of wild-type mice treated according to **a**. Each group comprises six wild-type mice. ‘†’ represents death of the mice. **c**, **d**, 50:50% wild-type:*Ifnar*^{-/-} (c) and 95:5% wild-type:*Ifnar*^{-/-} (d) bone marrow chimaeras (day 0) treated eight times with poly(I:C) (5 $\mu\text{g g}^{-1}$) every second day, analysed at day 22. Chimaerism in myeloid cells, lymphoid cells and HSCs in peripheral blood (PBL) or bone marrow HSCs (as indicated); CD45.1⁺ (filled bars), CD45.2⁺ (open bars). All data normalized to 5, 50 or 95%. Six mice were analysed per condition. Data are the mean and s.d. **e**, Model showing the activating effects of short-term (acute) IFN α stimulation on dormant/quiescent HSCs, and the inhibitory effects of chronic IFN α treatment on HSC self-renewal. IFN α binds to and activates IFNAR, resulting in phosphorylation of STAT1 and PKB. Subsequently, several IFN α target genes are expressed (Supplementary Table 1). Sca-1 is upregulated at the mRNA and protein level, and Sca-1 protein is incorporated into the plasma membrane. Although the signalling pathway downstream of Sca-1 remains unknown, it is required for IFN α -IFNAR-STAT1-induced proliferation because *Sca-1*^{-/-} mice do not respond to IFN α stimulation.

the possibility that IFN α pre-treatment in these patients might have activated and therefore sensitized the CML stem cells to imatinib mesylate. The so far unrecognized effects of acute and chronic IFN α signalling on stem cells may influence not only the future treatment of CML but potentially other diseases and malignancies as well.

METHODS SUMMARY

Mice. All mice were housed in individually ventilated cages in the EPFL/ISREC animal facility. Animal procedures were performed according to protocols approved by the Swiss Bundesamt für Veterinärwesen no. 1728. C57BL/6 mice are referred to as wild-type mice. *Ifnar*^{-/-}, *Sca-1*^{-/-} and *Stat1*^{-/-} mice are all on a C57BL/6 background²⁻⁴. CD45.1⁺ B6.SJL-Ptprca-Pep3b-/BoyJ donor mice were purchased from the Jackson Laboratory. Six-eight-week-old wild-type, *Ifnar*^{-/-}, *Sca-1*^{-/-} or *Stat1*^{-/-} mice were injected intraperitoneally (i.p.) with 5 μ g g⁻¹ poly(I:C) (Invivogen), or subcutaneously (s.c.) with 10,000 units per mouse of IFN α 4 (gift from D. Tough), unless indicated differently. Control mice were injected with PBS. For the LRC assay, mice were BrdU-labelled for 10 days using BrdU water (0.8 mg ml⁻¹, glucose) followed by a 100-day chase. 5-FU studies were performed using i.p. injections of 150 mg kg⁻¹ 5-fluoro-uracil (Sigma).

Preparation of bone marrow and PBL. To collect bone marrow cells, leg bones were crushed and cell suspensions were filtered before use. For FACS analysis of PBLs, six drops of blood were collected into a tube containing heparin, diluted in PBS, and centrifuged over a Histopaque-1083 (Sigma). PBLs were collected from the interface, washed and stained as described later.

FACS. Bone marrow cells were stained for haematopoietic subsets using lineage antibodies, cKit, Sca-1, CD135 (also known as Flt3), CD150 (Slamf1), CD34 and CD48. All cell suspensions were filtered through a nylon mesh filter (70 μ m) before FACS analysis to prevent clumping. Statistical analysis was performed using a two-tailed *t*-test. Statistical significance is indicated by **P* < 0.05 or ***P* < 0.01.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 September 2008; accepted 26 January 2009.

Published online 11 February 2009.

- Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
- Muller, U. *et al.* Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918–1921 (1994).
- Durbin, J. E., Hackenmiller, R., Simon, M. C. & Levy, D. E. Targeted disruption of the mouse *Stat1* gene results in compromised innate immunity to viral disease. *Cell* **84**, 443–450 (1996).
- Ito, C. Y., Li, C. Y., Bernstein, A., Dick, J. E. & Stanford, W. L. Hematopoietic stem cell and progenitor defects in *Sca-1*/Ly-6A-null mice. *Blood* **101**, 517–523 (2003).
- Randall, T. D. & Weissman, I. L. Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment. *Blood* **89**, 3596–3606 (1997).
- Kujawski, L. A. & Talpaz, M. The role of interferon- α in the treatment of chronic myeloid leukemia. *Cytokine Growth Factor Rev.* **18**, 459–471 (2007).
- Hellmann, R., Hochhaus, A. & Baccarani, M. Chronic myeloid leukaemia. *Lancet* **370**, 342–350 (2007).
- Trumpp, A. & Wiestler, O. D. Mechanisms of disease: cancer stem cells—targeting the evil twin. *Nature Clin. Pract. Oncol.* **5**, 337–347 (2008).
- Borden, E. C. *et al.* Interferons at age 50: past, current and future impact on biomedicine. *Nature Rev. Drug Discov.* **6**, 975–990 (2007).
- Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264 (1998).
- Darnell, J. E. Jr, Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415–1421 (1994).
- Pichlmair, A. & Reis e Sousa, C. Innate recognition of viruses. *Immunity* **27**, 370–383 (2007).
- Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
- Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Adolfsson, J. *et al.* Upregulation of Flt3 expression within the bone marrow Lin⁻Sca1⁺c-kit⁺ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15**, 659–669 (2001).
- Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
- Wilson, A. *et al.* Dormant and self-renewing hematopoietic stem cells and their niches. *Ann. NY Acad. Sci.* **1106**, 64–75 (2007).
- Wilson, A. & Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. *Nature Rev. Immunol.* **6**, 93–106 (2006).
- Yang, X. *et al.* Prostaglandin A₂-mediated stabilization of p21 mRNA through an ERK-dependent pathway requiring the RNA-binding protein HuR. *J. Biol. Chem.* **279**, 49298–49306 (2004).
- van Boxel-Dezaire, A. H., Rani, M. R. & Stark, G. R. Complex modulation of cell type-specific signaling in response to type I interferons. *Immunity* **25**, 361–372 (2006).
- Holmes, C. & Stanford, W. L. Concise review: stem cell antigen-1: expression, function, and enigma. *Stem Cells* **25**, 1339–1347 (2007).
- Lerner, C. & Harrison, D. E. 5-Fluorouracil spares hemopoietic stem cells responsible for long-term repopulation. *Exp. Hematol.* **18**, 114–118 (1990).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21^{clp1/waf1}. *Science* **287**, 1804–1808 (2000).
- Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nature Rev. Genet.* **9**, 115–128 (2008).
- O'Hare, T., Corbin, A. S. & Druker, B. J. Targeted CML therapy: controlling drug resistance, seeking cure. *Curr. Opin. Genet. Dev.* **16**, 92–99 (2006).
- Atallah, E. & Cortes, J. Optimal initial therapy for patients with newly diagnosed chronic myeloid leukemia in chronic phase. *Curr. Opin. Hematol.* **14**, 138–144 (2007).
- Heaney, N. B. & Holyoake, T. L. Therapeutic targets in chronic myeloid leukaemia. *Hematol. Oncol.* **25**, 66–75 (2007).
- Rousselot, P. *et al.* Imatinib mesylate discontinuation in patients with chronic myelogenous leukemia in complete molecular remission for more than 2 years. *Blood* **109**, 58–60 (2007).
- Hochhaus, A. First-line management of CML: a state of the art review. *J. Natl. Compr. Canc. Netw.* **6** (suppl. 2), S1–S10 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Aguet for discussions and advice throughout the project, and for providing mouse strains. We thank D. Tough for providing mouse recombinant IFN α 4, T. Pedrazzini, W. Stanford and M. Müller for mouse strains, K. Harshman and O. Hagenbüchle and the DAFL team for excellent service and help with the DNA microarrays. We thank C. Dubey and D. Aubry for animal husbandry, genetic screening and technical help, and J. Roberts for FACS sorting. We are grateful to A. Wilson for comments on the manuscript. M.A.G.E. is the recipient of an EMBO long-term fellowship. This work was supported by grants to A.T. from the Swiss National Science Foundation, the Swiss Cancer League, the EU-FP6 Program 'INTACT', the EU-FP7 Program 'EuroSyStem' and to UK from the Deutsche Forschungsgemeinschaft (SFB432.B15).

Author Contributions A.T., M.A.G.E. and U.K. designed the experiments and analysed the data. M.A.G.E., S.O. and Z.W. performed the experiments. W.E.B.-B. carried out the microarray analysis. A.T., M.A.G.E. and M.D. wrote the paper.

Author Information The microarray data have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE14361. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.T. (a.trumpp@dkfz-heidelberg.de).

METHODS

ELISA for IFN α . Wild-type mice were injected with the indicated amounts of poly(I:C). After 5 h, blood samples were collected and the concentration of mouse IFN α in the blood serum was determined by an ELISA kit according to the manufacturer's instructions (PBL Biomedical Laboratories).

Generation and analysis of chimaeras. CD45.1⁺ B6.SJL-Ptprca-Pep3b-/BoyJ (the Jackson Laboratory) donor mice were purchased and maintained in the ISREC animal facility. To generate mixed chimaeras, transplantations were performed using mixtures of CD45.1⁺ wild-type bone marrow and CD45.2⁺ *Ifnar*^{-/-} bone marrow with a total amount of 3×10^7 bone marrow cells (Fig. 2a). Bone marrow mixtures were injected intravenously (i.v.) into lethally irradiated CD45.1⁺ wild-type recipient mice that had been pre-treated (48 h) with anti-NK1.1 monoclonal antibody. All chimaeric mice were maintained on antibiotic containing water (Bactrim, Roche). To generate reverse chimaeras, 3×10^7 CD45.1⁺ wild-type bone marrow cells were injected i.v. into lethally irradiated CD45.2⁺ *Ifnar*^{-/-} mice, and bone marrow cells were analysed 1–2 months later. Three months after the bone marrow transplantation, mice were injected with IFN α and analysed for HSC proliferation.

Long-term reconstitution assay. Twenty-two days after the start of the chronic IFN α treatment, bone marrow from treated 50:50% wild-type:*Ifnar*^{-/-} and 95:5% wild-type:*Ifnar*^{-/-} chimaeras was isolated. Then, 3×10^7 bone marrow cells were injected i.v. into lethally irradiated CD45.1⁺ wild-type recipient mice. Starting at week 6, reconstitution of PBL was analysed by tail vein bleeding every second week.

For the long-term reconstitution assay of wild-type versus *Ifnar*^{-/-} bone marrow, and wild-type versus three-times poly(I:C)-injected bone marrow, 50,000 bone marrow cells from either wild-type (CD45.2⁺) or *Ifnar*^{-/-} (CD45.2⁺) mice were together with 500,000 'rescue' bone marrow cells (Sca1-depleted CD45.1⁺ bone marrow cells) injected i.v. into lethally irradiated CD45.1⁺ wild-type recipient mice.

Sorting of HSCs. To isolate HSCs, lineage magnetic depletion was performed to enrich for lineage-negative cells. Bone marrow cells were incubated with lineage antibodies (CD4, CD8, CD11b, Gr1, B220 and Ter119) and lineage-positive cells were removed using sheep anti-rat IgG-coated M450 Dynalbeads (Dynal Biotech). Lin^{neg} cells were stained for haematopoietic subsets and LSK-HSCs were sorted using a BD FACS Vantage-DIVA (Becton Dickinson).

Phosflow-staining FACS. For staining of pSTAT1 or pPKB, LSK-HSCs were sorted as described above, and stained according to the BD phosflow (Becton Dickinson) protocol for mouse cells using the STAT1 (pY701) (clone 4a)-Alexa647 or Akt (pS473)-Alexa647 antibody from BD phosflow.

BrdU staining and cell-cycle analysis. Proliferation analyses on HSCs were performed using cell surface staining in combination with BrdU. For BrdU analysis, mice were injected i.p. with BrdU (7.2 mg kg⁻¹, Sigma) 12 h before analysis, unless indicated differently. Mice were killed and bone marrow cells were isolated. Bone marrow cells were stained for haematopoietic subsets, and for BrdU staining a commercially available kit (BD Biosciences) was used.

Cell cycle analysis on HSCs was performed using cell surface staining in combination with intracellular Ki67 and Hoechst staining. In brief, bone marrow cells were labelled with monoclonal antibody conjugates to surface markers,

fixed in 2% paraformaldehyde in PBS, washed and stained with anti-Ki67-FITC (BD Biosciences) for 30 min. During the last 10 min, cells were co-stained with Hoechst 33342 (Molecular Probes) at 20 μ g ml⁻¹.

FACS. For flow cytometric analysis the 6 colour BD FACS Canto (BD Biosciences) equipped with a 488 nm and a 633 nm laser, the 8 colour BD FACS Vantage-DIVA (BD Biosciences) equipped with a 488 nm laser, a Multiline UV and a 647 nm laser, or the 9 colour analyser Cyan ADP Analyser (Beckman Coulter) equipped with a 488 nm, a 635 nm and 407 nm laser were used.

Antibodies. Gr-1 (Ly-6G, RB6-8C5)-FITC, -biotin and -Alexa647, Ter119-FITC and -biotin, B220 (RA3-6B2)-FITC and -biotin, CD11b-FITC and -biotin, CD4 (clone GK1.5)-FITC and -biotin, CD8 α (53.6.7)-FITC and -biotin, CD45.1 (A20.1)-FITC, -biotin, -phycoerythrin (PE) or -Alexa647, CD45.2 (ALI-4A2)-FITC, -biotin or -Alexa647 were purified and conjugated in this laboratory following standard protocols. CD34 (RAM34)-FITC, CD135 (A2F10)-PE, cKit (2B8)-PE, -PeCy5, -PeCy7, -APC, and -APCCy7, Sca-1 (D7)-APC and (2B8)-biotin, CD48 (HM48-1)-PE, Sca1 (D7)-APC, and -Alexa700, CD4 (GK1.5)-PeCy7, CD8 (53-6.7)-PeCy7, CD11b (M1/70)-PeCy7, Gr1 (RB6-8C5)-PeCy7, B220 (RA3-6B2)-PeCy7, and Ter119-PeCy7 were purchased all from eBioscience. CD150 (TC15-12F12.2)-PeCy5 was purchased from Biolegend.

RNA isolation, microarray amplification and hybridization methods. Total RNA was isolated from sorted Lin^{neg} cKit⁺ CD150⁺ CD48⁻ cells using the Qiagen RNA isolation kit (Qiagen). Per condition, three independent samples were analysed. Two rounds of amplification for each RNA sample were performed using the Nugen WT-Ovation Pico RNA Amplification System (Nugen). Biotin labelling of cRNA was performed using the Affymetrix GeneChip IVT labelling kit (Affymetrix). This biotinylated RNA was fragmented and hybridized to Affymetrix MOE430v2 chips (Affymetrix) as per the manufacturer's protocol. Both raw image (.dat) and intensity (.cel) files were generated using the Affymetrix Gene Chip Operating Software.

Microarray analysis. Quality control tests were performed using the DNA Array Facility of Lausanne's Remote Analysis System (<http://race.unil.ch>). This is a web-based interface for various statistical analysis routines performed using the R language (<http://www.r-project.org>). These involved various quality tests such as the comparison of RNA 5'-end to 3'-end bias using RNA degradation plots to determine the quality of amplification, density of perfect match (PM) intensities, RMA normalization of the chips, sample clustering to control replicates, correlation matrix to observe correlation between samples and replicates, surface intensity in log scale to examine for chip and hybridization defects, scatter plots to compare samples and replicates. Chips that did not fit these quality checks were eliminated from the sample and not considered in the following statistical analysis. RMA-normalized data was entered into the Genespring program (Silicon Genetics) for data visualization as well as further filtering and examination of overlaps of various gene lists.

Samples having passed the quality check were then further statistically analysed using the Statistical Bayes Test Module on the RACE system. Genes showing a twofold difference with a minimum Bayesian *P* value of 0.05 were then selected. These genes were then further analysed for significant genes and pathways using Ingenuity Pathways Analysis (Ingenuity Systems).

Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site

Puwei Yuan^{1*}, Mark Bartlam^{2*}, Zhiyong Lou^{3*}, Shoudeng Chen¹, Jie Zhou¹, Xiaojing He¹, Zongyang Lv¹, Ruowen Ge⁴, Xuemei Li^{1,3}, Tao Deng^{2,5}, Ervin Fodor⁵, Zihao Rao^{1,2,3} & Yingfang Liu¹

The heterotrimeric influenza virus polymerase, containing the PA, PB1 and PB2 proteins, catalyses viral RNA replication and transcription in the nucleus of infected cells. PB1 holds the polymerase active site¹ and reportedly harbours endonuclease activity², whereas PB2 is responsible for cap binding^{2–4}. The PA amino terminus is understood to be the major functional part of the PA protein and has been implicated in several roles, including endonuclease⁵ and protease activities⁶ as well as viral RNA/complementary RNA promoter binding⁷. Here we report the 2.2 Å crystal structure of the N-terminal 197 residues of PA, termed PA_N, from an avian influenza H5N1 virus. The PA_N structure has an α/β architecture and reveals a bound magnesium ion coordinated by a motif similar to the (P)DX_N(D/E)XK motif characteristic of many endonucleases. Structural comparisons and mutagenesis analysis of the motif identified in PA_N provide further evidence that PA_N holds an endonuclease active site. Furthermore, functional analysis with *in vivo* ribonucleoprotein reconstitution and direct *in vitro* endonuclease assays strongly suggest that PA_N holds the endonuclease active site and has critical roles in endonuclease activity of the influenza virus polymerase, rather than PB1. The high conservation of this endonuclease active site among influenza strains indicates that PA_N is an important target for the design of new anti-influenza therapeutics.

Highly pathogenic avian influenza A virus strains with H5N1 subtype are entrenched in poultry worldwide and pose a growing threat to human health. Of the 387 reported human cases of avian influenza since 2003, 245 have been fatal (World Health Organisation, September 2008). Understanding the viral replication mechanism mediated by the viral RNA polymerase is crucial for the development of new anti-influenza therapeutics to increase preparedness against a global influenza pandemic.

The influenza virus RNA-dependent RNA polymerase is a heterotrimeric complex (PA, PB1 and PB2) harbouring several enzymatic activities for catalysing both viral RNA transcription (vRNA \rightarrow messenger RNA) and replication (vRNA \rightarrow cRNA \rightarrow vRNA). The functions of PB1 and PB2 are well defined², whereas PA has been implicated in a diverse range of functions^{5,8–15}. Mutational analysis of PA suggests that PA_N is involved in several functions of the polymerase complex, including protein stability, endonuclease activity, cap binding and promoter binding^{5,7}.

The PA protein can be cleaved by limited tryptic digestion into two domains: a smaller N-terminal ~ 25 kDa domain, and a larger carboxy terminal ~ 55 kDa domain^{5,16}. The crystal structure of the large C-terminal domain of PA (residues 257–716, termed PA_C), understood to bind to PB1 for complex formation and nuclear transport, was recently determined by our group and others^{17,18}. The structure of

PA_C in complex with an inhibitory N-terminal peptide from PB1 (termed PB1_N) provides a starting point for further investigation into the structure and function of the influenza virus polymerase and a target for the discovery of anti-influenza therapeutics. In particular, the high conservation of the PB1_N binding site points towards a viable target to design alternatives to the influenza drugs oseltamivir (Tamiflu) and zanamivir (Relenza), which are at present beset by problems of resistance. Here we extend our previous study on the C-terminal region of PA by investigating the structure and function of the smaller PA N-terminal domain.

The N-terminal domain of PA covering residues 1–256 and termed PA_N, was cloned from an avian type A virus isolate (A/goose/Guangdong/1/96 (H5N1)) and expressed in *Escherichia coli*. Recombinant PA_N was purified and crystallized, and the subsequent crystal structure of PA_N was determined by multi-wavelength anomalous dispersion using 2.9 Å data from a seleno-methionyl derivative protein and 2.2 Å native data (Fig. 1a). From the resulting experimental electron density map, PA_N was traced from residues 1–197, with the exception of a flexible loop region from residues 49–75 and the entire C-terminal region from residues 198–256. Data collection, phasing and refinement statistics are listed in Supplementary Table 1.

The PA_N structure has an α/β architecture with five mixed β -strands forming a twisted plane surrounded by seven α -helices (Fig. 1a, b). A negatively charged cavity surrounded by helices $\alpha 2$ – $\alpha 5$ and strand $\beta 3$ houses a bound metal ion (Fig. 1c), which we identified as magnesium for the following reasons: a high concentration of magnesium (0.1 M) was added to the crystallization solution, and atomic absorption results showed a significant amount of magnesium in the washed and resolved PA_N crystals in an approximately 13:9 molar ratio (data not shown). The metal is directly coordinated by five ligands: the acidic residues E80 and D108, three water molecules stabilized by residue H41, E119 and the carbonyl oxygens of L106 and P107 (Fig. 1d). All six amino acids involved in coordinating Mg²⁺ are conserved among influenza A, B and C viruses, with the exception of P107 which is replaced with alanine and cysteine in influenza B and C viruses, respectively (Supplementary Fig. 1).

A Dali (<http://www.ebi.ac.uk/dali>) search identified several similar structures to PA_N, including: the Tt1808 hypothetical protein from *Thermus thermophilus* HB8 (Protein Data Bank (PDB) accession 1WDJ, Z-score 4.8, root mean squared deviation (r.m.s.d.) 3.4 Å for 87 amino acids), the restriction endonuclease SdaI (PDB accession 2IXS, Z-score 3.9, r.m.s.d. 4.0 Å for 95 amino acids) and the Holliday junction resolvase Hjc (PDB accession 1GEF, Z-score 3.8, r.m.s.d. 3.0 Å for 76 amino acids). SdaI belongs to the type IIP DNA restriction endonucleases and its C-terminal domain exhibits a classic nuclease motif containing a seven-stranded planar β -sheet surrounded by four

¹National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ²College of Life Sciences and Tianjin Key Laboratory of Protein Science, Nankai University, Tianjin 300071, China. ³Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China. ⁴Department of Biological Sciences, National University of Singapore, Singapore 117543. ⁵Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK.

*These authors contributed equally to this work.

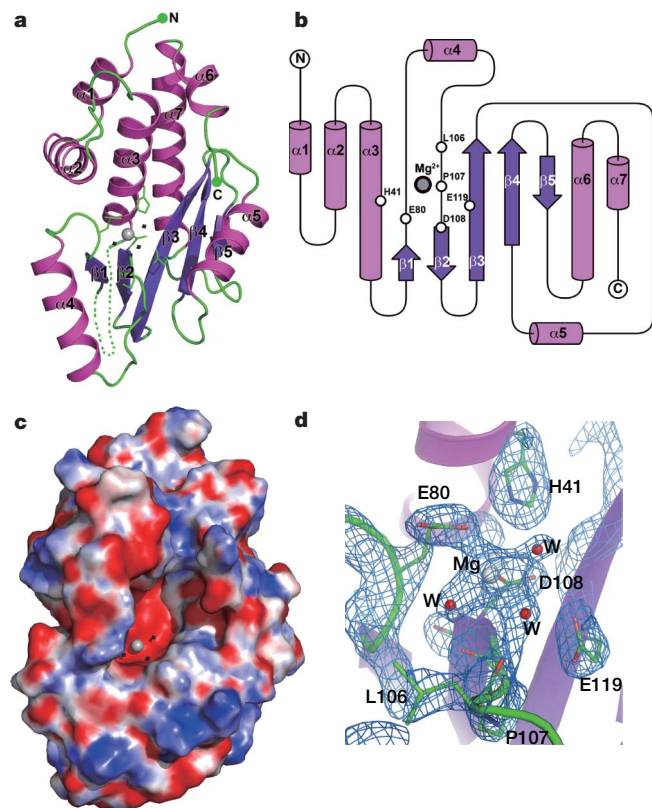


Figure 1 | The PA_N structure. **a**, Ribbon representation showing the PA_N structure. The structure is coloured according to secondary structure elements: α -helices are pink, β -strands are magenta, and loops are green. Individual secondary structure elements are labelled. The Mg²⁺ ion is shown by a silver sphere and the three water molecules are indicated by black dots. **b**, Topology figure of the PA_N structure coloured according to the scheme in **a**. **c**, Surface representation showing the same view of PA_N as in **a**, coloured by electrostatic charge from red ($-10 k_B T/e_c$, in which k_B is the Boltzmann constant, T is temperature and e_c is the electron charge) to blue ($+10 k_B T/e_c$). The Mg²⁺ ion is shown as a silver sphere and water molecules are shown by black spheres. **d**, Close-up view of the Mg²⁺ binding site covered by a $2F_o - F_c$ electron density map (contoured at 1.5σ). Residues coordinating the Mg²⁺ ion are shown in stick representation and labelled. The Mg²⁺ ion is shown by a silver sphere and water molecules are shown by red spheres. The PA_N structure is in the same orientation and coloured according to the scheme in **a**.

α -helices. The hypothetical Tt1808 is also predicted to be an endonuclease on the basis of structural similarity, and both proteins contain a conserved (P)DX_N(D/E)XK active site motif characteristic of type II endonucleases, with only the lysine position in the motifs varying. The overall structure of PA_N is comparable to both SdaI and Tt1808 (Fig. 2a, b). In the case of SdaI, PA_N shares a similar core structure and strands β 1– β 5 match closely with the central β -sheet (strands β 8 and β 10– β 13) of SdaI (Fig. 2a). There is particular similarity between residues coordinating Mg²⁺ in PA_N and residues in the SdaI active site motif. P107, D108 and E119 of PA_N align well to P232, D233 and E248 of SdaI, respectively, suggesting that PA_N might contain endonuclease activity with a motif similar to the (P)DX_N(D/E)XK motif characteristic of many endonucleases (Fig. 2c, d).

The endonuclease activity of the influenza virus polymerase is critical for snatching capped primers from host mRNA to initiate mRNA transcription. PA is reported to be involved in endonuclease activity, either via its N- or C-terminal domain^{5,11}; in particular, several mutations in the N-terminal of PA inhibited this activity⁵. However, it was previously reported that the endonuclease activity of the influenza polymerase is located in PB1, and the acidic residues E508, E519 and D522 are essential for this activity², whereas earlier studies suggested that the PB2 subunit might contain the endonuclease active site^{19–21}.

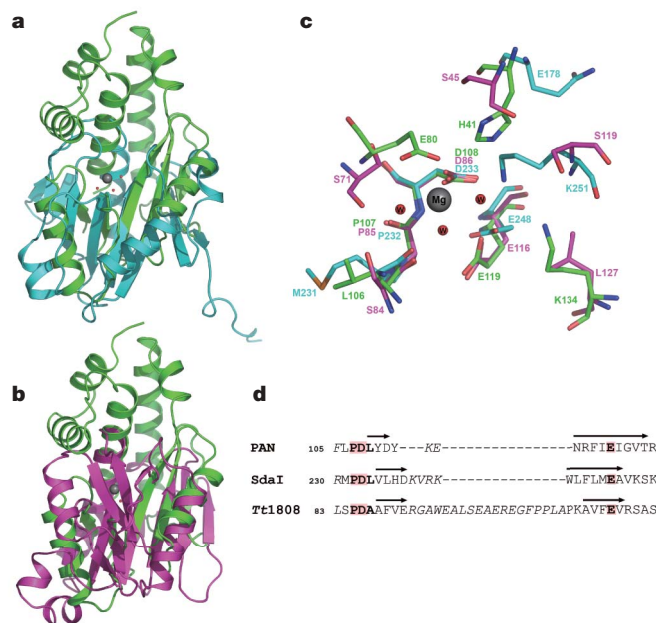


Figure 2 | Structural comparisons suggest PA_N holds an endonuclease active site. **a**, Superposition of the PA_N (green) and SdaI (cyan) structures, viewed in the same orientation as Fig. 1a. **b**, Superposition of the PA_N (green) and Tt1808 (magenta) structures, viewed in the same orientation as Fig. 1a. **c**, Superposition of the putative active site residues of PA_N (green) with the active sites of SdaI (cyan) and Tt1808 (magenta). Residues are shown in stick representation and labelled. The Mg²⁺ ion and coordinating water molecules from the PA_N structure are shown as silver and red spheres, respectively. **d**, Structure-based sequence alignment of PA_N with SdaI (PDB accession 2IXS) and Tt1808 (PDB accession 1WDJ), showing the structurally-conserved endonuclease motif (highlighted in red). Amino acids shown in italic are not aligned.

To resolve this issue, we mutated residues in PA_N involved in metal coordination, including H41, E80, L106, P107, D108 and E119, and performed *in vivo* ribonucleoprotein (RNP) reconstitution assays followed by primer extension analysis of the steady-state levels of all three species of viral RNAs synthesized by the virus polymerase (Fig. 3a)¹¹. We also included residue K134, which is located close to the proposed endonuclease active site and has been shown to specifically inhibit viral RNA transcription but not replication in previous studies⁵. Polymerases with E80A, D108, E119A and K134A point mutations in PA showed background levels of mRNA synthesis, while retaining significant cRNA and vRNA synthesis activity, in comparison with the wild-type polymerase (Fig. 3a, b). In contrast, L106A and P107A retained mRNA synthesis activity (L106A at near wild-type levels), probably because they interact with the Mg²⁺ ion via their carbonyl oxygens (Figs 1d and 3a). However, they both showed a reduction in cRNA and vRNA synthesis, whereas H41A showed no detectable synthesis of any of the three viral RNAs. Notably, point mutations of the PB1 residues, E508, E519 and D522 (Fig. 3a), which have previously been claimed to be the polymerase endonuclease active centre², resulted in significant levels of activity. Sequence similarity searches do not identify a possible endonuclease activity motif around residues E508, E519 and D522 in PB1 (results not shown). These observations strongly suggest that PA provides a centre for polymerase endonuclease activity, rather than PB1.

To address this further, we partially purified a recombinant viral RNA polymerase complex with mutations in the putative endonuclease active site from transfected 293T cells (Fig. 3c). Approximately equal amounts of RNA polymerase were used in transcription assays *in vitro* using an ApG primer that circumvents the need for cap-binding and endonuclease cleavage for transcription initiation. Wild-type or near wild-type levels of activity were observed for the E80A, D108A,

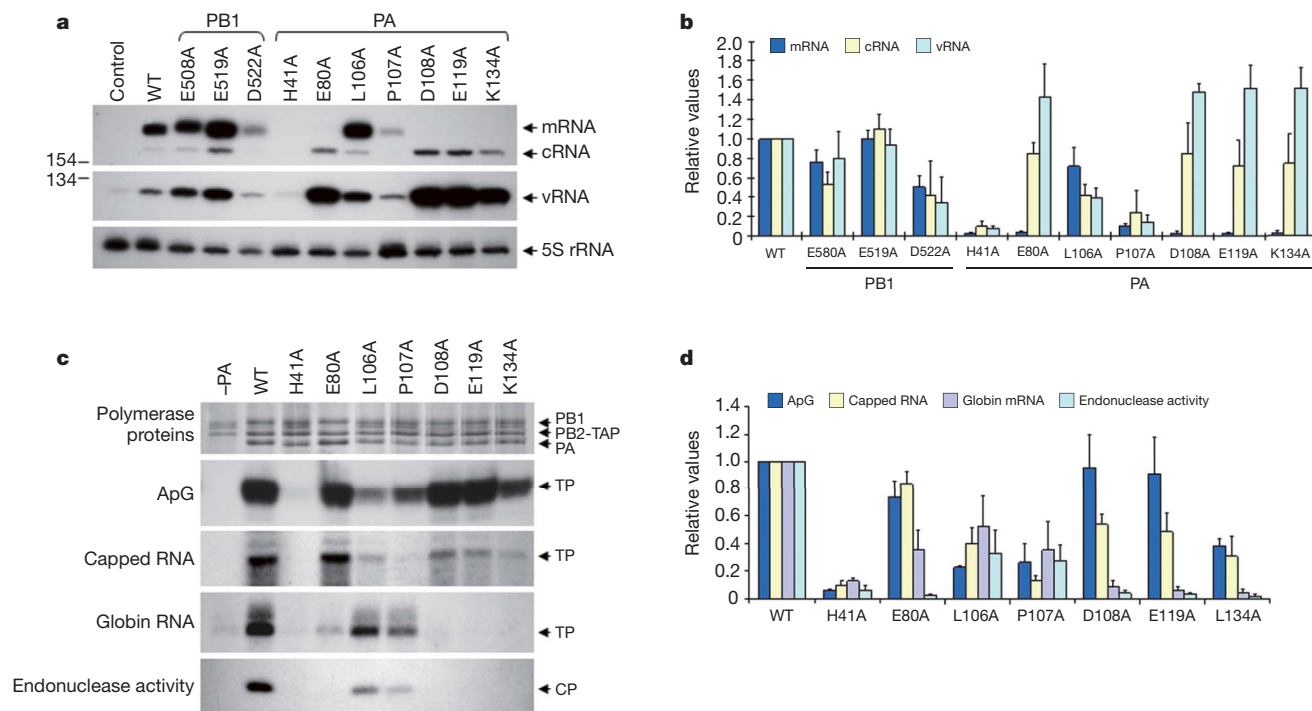


Figure 3 | Effects of mutations in the PA and PB1 subunits on RNA polymerase function. **a**, The effect of polymerase mutations on viral RNA transcription and replication. Wild-type or mutant ribonucleoprotein complexes were reconstituted in 293T cells and viral RNA levels were analysed using a primer extension assay. **b**, Quantification of viral RNA levels from **a** by phosphorimaging. **c**, Effects of mutations on polymerase complex formation, ApG, capped RNA and globin mRNA primed

E119A and K134A mutants, whereas the L106A and P107A mutants showed reduced but detectable activity (Fig. 3c, d). No activity was detected for the H41A mutant. Similar results were observed if a capped RNA primer, requiring no endonuclease cleavage before elongation, was provided. In contrast, if globin mRNA was provided as a source of capped RNA primer, requiring endonuclease cleavage, most mutations with the exception of L106A and P107A resulted in diminished activity, strongly suggesting that the ability of the polymerase to endonucleolytically cleave capped RNA is primarily affected by the mutations. Indeed, a direct endonuclease cleavage assay using the mutant polymerases showed that all mutants had no detectable activity, with the exception of the L106A and P107A mutations that are proposed to be involved via their carbonyl oxygens (see earlier). It should be noted that although the E80A mutant exhibited some activity in globin mRNA primed transcription, no activity was detected in the endonuclease assay; this may be owing to the different nature of the capped RNA used in the different assays (globin mRNA versus poly(A)-tailed capped RNA). These results obtained *in vitro* are in agreement with the results obtained in the cell-based RNP reconstitution assay and fully support the involvement of this region of PA in endonuclease cleavage activity. Nuclease contamination of PA_N purified from *E. coli* made it impossible to determine whether the isolated PA_N domain has endonuclease activity *in vitro* (data not shown).

Mutation of several residues in the region from D108–F117 of PA_N, including W88, P103, Y110 and F117, results in impaired polymerase activity affecting both transcription and replication⁵, whereas the D111A/Y112A double mutation reportedly interferes with transcription and replication⁹. W88, P103, Y110 and F117 form part of a hydrophobic region linking helix α 4 to the rest of the PA_N structure (Supplementary Fig. 2a), whereas D111 and Y112 interact with R75 on the loop between α 3 and β 1 (Supplementary Fig. 2b). Their

transcription initiation, and endonuclease activity. The position of the three polymerase subunits, PB1, PB2-TAP and PA is indicated. The position of the transcription products (TP) and endonuclease cleavage products (CP) is also indicated. The major, 7-nucleotide capped cleavage product is shown. **d**, Quantification of the results from **c** obtained by phosphorimaging analysis. Data presented in **b** and **d** are an average of three independent experiments, and standard deviations are shown.

respective mutations would be expected to destabilize the protein structure and therefore impair its function. Hence the region covering D108 to F117, which overlaps with the endonuclease active site motif, should also be important for the correct folding and stability of the protein and therefore essential for its function. Furthermore, our results indicate that the H41A, L106A and P107A mutations affected the replicative activity of the viral RNA polymerase, suggesting that Mg²⁺ binding might also contribute to the structural integrity of the polymerase in addition to its required role for endonuclease cleavage of capped RNA.

Taken together, the structure and functional assays demonstrate that PA_N contains an endonuclease active site with a putative P₁₀₇D₁₀₈X₁₀₉E₁₁₉K₁₃₄ active site motif. The first step in the general mechanism of phosphodiester hydrolysis is the preparation of the attacking nucleophile by deprotonation, usually involving a base to deprotonate a water molecule. Lysine is often considered as a general base candidate in endonucleases but it is not strictly conserved^{22,23}. The prime candidate for this lysine in PA_N is K134, which is oriented towards the active site; a K134A mutant in functional assays shows a phenotype fully compatible with being involved in endonucleolytic activity. However, from the structure, K134 is located about 7 Å from the Mg²⁺ ion and is therefore too far from the active site to be directly involved. We speculate that in the context of the polymerase trimer, the relative position of α 5 harbouring K134 might be slightly altered allowing its direct involvement. A second possibility is that an alternative residue, that is, H41, might serve as the base and K134 has an indirect role.

Our structural data provide no information on how the capped RNA substrate could access the endonuclease active site as it reveals no obvious sites for RNA binding. One possible RNA binding site could be formed by a cluster of four arginines on the protein surface:

two arginine residues (R124 and R125) on the $\beta 3$ – $\alpha 5$ loop, and two arginines (R192 and R196) on helix $\alpha 7$ (Supplementary Fig. 2c). Capped RNA has to be accessed by PB2 and PB1, with PB2 providing the cap-binding domain and PB1 being involved in the elongation of cleaved capped RNA. Previous studies have shown that capped RNA can be cross-linked to PB2 and PB1 (refs 2, 4), but no direct interaction has been reported for PA and capped RNA. We speculate that the capped RNA binding site might be formed by residues contributing from PB1 and PB2 and a co-crystal structure of PA with capped RNA, and possibly including PB1 and PB2, might be required to resolve this issue. PA_N is also associated with the nuclear transport of PB1 (refs 24, 25), and the arginine cluster mentioned above coincides with two putative nuclear transport motifs that are proposed to lie within the N-terminal domain of PA: residues 124–139 (NLS1) and 186–247 (NLS2)²⁶. This arginine cluster exposed on the protein surface is consistent with the previous proposal²⁶ that PA_N contains a bipartite nuclear localization signal.

PA has been linked to proteolysis of viral and host proteins. Residues 1–247 are claimed to be sufficient to induce protein degradation *in vivo* with T157 being involved in both protease and polymerase replication activities^{6,27}, whereas a conflicting report places S624 of PA_C at the protease active site¹². Inspection of the environment surrounding T157 in the PA_N structure reveals no obvious candidates for a protease active site (Supplementary Fig. 2d), and our own *in vitro* protease assays following the protocol previously described¹² indicate that PA_N has no detectable proteolytic activity (Supplementary Fig. 3). Further studies are therefore required to clarify the role of PA in protease activity and to determine the location of the active site. Nevertheless, several residues surrounding T157, including E154, K158, D160, E165, E166, R168 and R170, are highly conserved across influenza species (Supplementary Fig. 1), suggesting that this region is an important part of the polymerase complex. For example, region 163–178 of PA has been implicated in the regulation of the promoter binding activities of the RNA polymerase⁷.

The availability of structures for the N- and C-terminal domains of PA enables us to build up a more complete picture of the PA protein (Supplementary Fig. 4). PA_N was expressed from residues 1–256, although residues 198–256 could not be observed from the structure, and PA_C expressed from residues 257–716 lacks any regular secondary structure in its N-terminal 17. This leads us to suspect that the region linking PA_N and PA_C is flexible. Our attempts to express the central linker region of PA from residues 197–256 alone proved unsuccessful, and secondary structure predictions indicate that this region should possess more than 60% random coil (data not shown). A recent study confirmed our observations and showed that neither the N- nor the C-terminal domains of PA could ensure a stable interaction with PB1 without the presence of the linker¹⁶. Furthermore, protease treatment of the PA–PB1 complex showed that its PA protein is markedly more stable than free PA, indicating that the linker is protected from digestion by PB1 and forms an essential part of the subunit interface¹⁶. This interface spans much of the PA sequence, with the exception of the N-terminal 154 amino acids, and is consistent with previous reports mapping the interaction domain of PA and PB1 to the C-terminal two-thirds of PA^{13,28}. The presence of the linker between the N- and C-terminal domains of PA provides a degree of conformational flexibility that may enable it to have a role in regulating polymerase functions through conformational changes of polymerase complexes¹⁵. In fact, mutations in the linker region are known to affect polymerase function: L226P results in accumulation of the polymerase complex in an inactive form, whereas F205A and L214A mutants diminish RNA synthesis activity⁵.

We have determined the crystal structure of the N-terminal domain of PA from residues 1–197 and identified an endonuclease active motif in this domain. Mutational analysis of the identified motif confirmed that PA_N, rather than PB1, harbours an endonuclease active site and suggests that the activity should be Mg²⁺-dependent. Furthermore,

the high conservation of residues in the site indicates that it is general to all PA_N of different influenza viruses. This provides a feasible target for the discovery of new anti-influenza therapeutics, which are urgently needed owing to the increasing problems of resistance to available influenza drugs such as oseltamivir (Tamiflu) and zanamivir (Relenza). The PA_N structure reported here, combined with our previous structure of the C-terminal domain of PA, offers a more complete view of the PA protein and its functional roles in the influenza polymerase complex. In particular, our work provides a much-needed structural basis for researchers to understand the role of PA_N in protein stability, endonuclease activity, cap binding and vRNA promoter binding.

METHODS SUMMARY

Residues 1–256 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and overexpressed in *E. coli* strain BL21. The recombinant protein was purified with a glutathione affinity column (GE Healthcare). Glutathione S-transferase (GST) was cleaved with PreScission protease (GE Healthcare). The protein complex was further purified by Q ion exchange and Superdex-200 gel filtration chromatography (GE Healthcare).

Native PA_N was crystallized in the space group *P1* using 25% PEG8000 at pH 6.5. A selenomethionyl derivative was crystallized in *P6*₄22 using 20% PEG3350 at pH 6.5. The structure was phased to 2.9 Å by multiple-wavelength anomalous dispersion from a selenomethionyl derivative, and traced using 2.2 Å native data. The final model, with an *R*-factor of 23% and *R*_{free} of 25%, contains residues 1–197 of PA and is missing residues 49–75.

Plasmids pcDNA-PA, pcDNA-PB1, pcDNA-PB2, pcDNA-NP and pPOLI-NA-RT used in ribonucleoprotein reconstitution assays were identical to those in previous reports^{11,29}. Primer extension assays of viral RNAs were performed as previously described²⁹.

Point mutations were introduced using PCR by designing mutated residues in primers. Mutated genes were cloned into the pcDNA3.1 vector. The pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-TAP plasmids were co-transfected into 293T cells, which were collected 2 days later. Recombinant proteins were purified by tandem affinity purification (TAP) and stored at –80 °C⁵. *In vitro* transcription assays using ApG or capped RNA primers or globin mRNA as a source of capped RNA primer were performed as described^{8,11}. Capped RNA (m⁷GpppACACUUGCUUUUG) for the capped RNA primed transcription assay has been described³⁰. The capped ³²P-labelled RNA was poly(A)-tailed for endonuclease cleavage assays¹¹. A 7-nucleotide major and a 13-nucleotide minor cleavage product were identified in endonuclease cleavage reactions, in agreement with the finding that endonucleolytic cleavage occurs preferentially at purine residues.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 November; accepted 12 December 2008.

Published online 4 February 2009.

- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874 (1989).
- Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20**, 2078–2086 (2001).
- Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15**, 500–506 (2008).
- Fechter, P. *et al.* Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J. Biol. Chem.* **278**, 20381–20388 (2003).
- Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80**, 7789–7798 (2006).
- Sanz-Ezquerro, J. J., Zurcher, T., de la Luna, S., Ortin, J. & Nieto, A. The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *J. Virol.* **70**, 1905–1911 (1996).
- Maier, H. J., Kashiwagi, T., Hara, K. & Brownlee, G. G. Differential role of the influenza A virus polymerase PA subunit for vRNA and cRNA promoter binding. *Virology* **370**, 194–204 (2008).
- Lee, M. T. *et al.* Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic Acids Res.* **30**, 429–438 (2002).
- Regan, J. F., Liang, Y. & Parslow, T. G. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J. Virol.* **80**, 252–261 (2006).

10. Naffakh, N., Massin, P. & van der Werf, S. The transcription/replication activity of the polymerase of influenza A viruses is not correlated with the level of proteolysis induced by the PA subunit. *Virology* **285**, 244–252 (2001).
11. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76**, 8989–9001 (2002).
12. Hara, K. *et al.* Influenza virus RNA polymerase PA subunit is a novel serine protease with Ser624 at the active site. *Genes Cells* **6**, 87–97 (2001).
13. Zurcher, T., de la Luna, S., Sanz-Ezquerro, J. J., Nieto, A. & Ortin, J. Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J. Gen. Virol.* **77**, 1745–1749 (1996).
14. Huarte, M. *et al.* Threonine 157 of influenza virus PA polymerase subunit modulates RNA replication in infectious viruses. *J. Virol.* **77**, 6007–6013 (2003).
15. Kawaguchi, A., Naito, T. & Nagata, K. Involvement of influenza virus PA subunit in assembly of functional RNA polymerase complexes. *J. Virol.* **79**, 732–744 (2005).
16. Guu, T. S., Dong, L., Wittung-Stafshede, P. & Tao, Y. J. Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. *Virology* **379**, 135–142 (2008).
17. He, X. *et al.* Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454**, 1123–1126 (2008).
18. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454**, 1127–1131 (2008).
19. Blok, V. *et al.* Inhibition of the influenza virus RNA-dependent RNA polymerase by antisera directed against the carboxy-terminal region of the PB2 subunit. *J. Gen. Virol.* **77**, 1025–1033 (1996).
20. Honda, A., Mizumoto, K. & Ishihama, A. Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl Acad. Sci. USA* **99**, 13166–13171 (2002).
21. Shi, L., Summers, D. F., Peng, Q. & Galarz, J. M. Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. *Virology* **208**, 38–47 (1995).
22. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature* **368**, 660–664 (1994).
23. Lukacs, C. M., Kucera, R., Schildkraut, I. & Aggarwal, A. K. Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 Å resolution. *Nature Struct. Biol.* **7**, 134–140 (2000).
24. Fodor, E. & Smith, M. The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J. Virol.* **78**, 9144–9153 (2004).
25. Nieto, A. *et al.* Nuclear transport of influenza virus polymerase PA protein. *Virus Res.* **24**, 65–75 (1992).
26. Nieto, A., de la Luna, S., Barcena, J., Portela, A. & Ortin, J. Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *J. Gen. Virol.* **75**, 29–36 (1994).
27. Perales, B. *et al.* The replication activity of influenza virus polymerase is linked to the capacity of the PA subunit to induce proteolysis. *J. Virol.* **74**, 1307–1312 (2000).
28. Toyoda, T., Adyshev, D. M., Kobayashi, M., Iwata, A. & Ishihama, A. Molecular assembly of the influenza virus RNA polymerase: determination of the subunit-subunit contact sites. *J. Gen. Virol.* **77**, 2149–2157 (1996).
29. Vreede, F. T., Jung, T. E. & Brownlee, G. G. Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *J. Virol.* **78**, 9568–9572 (2004).
30. Brownlee, G. G., Fodor, E., Pritlove, D. C., Gould, K. G. & Dalluge, J. J. Solid phase synthesis of 5'-diphosphorylated oligoribonucleotides and their conversion to capped m7Gppp-oligoribonucleotides for use as primers for influenza A virus RNA polymerase *in vitro*. *Nucleic Acids Res.* **23**, 2641–2647 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Chen and K. Yu for providing the A/goose/Guangdong/1/96 influenza PA gene, R. Zhang and A. Joachimiak for assistance with data collection, and G. G. Brownlee for providing materials and for valuable advice. This work was supported by the National Natural Science Foundation of China (grant numbers 30599432, 30221003 and 30721003), the Ministry of Science and Technology International Cooperation Project (2006DFB32420), the Ministry of Science and Technology 863 Project (2006AA02A314 and 2006AA02A322), the Ministry of Science and Technology 973 Project (2006CB504300 and 2007CB914300) and the Medical Research Council (G0700848).

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under the accession number 3EBJ. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.R. (raozh@xtal.tsinghua.edu.cn) and Y.L. (liuy@ibp.ac.cn).

METHODS

Protein expression and purification. Residues 1–256 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96 (H5N1)) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and overexpressed in *E. coli* strain BL21. The recombinant proteins were then purified with a glutathione affinity column (GE Healthcare). After cleavage of the GST tag with PreScission protease (GE Healthcare), the protein complex was further purified by Q ion exchange chromatography and then by Superdex-200 gel filtration chromatography (GE Healthcare).

Data collection and structure determination. Crystals were obtained by the vapour diffusion method and grown using with 25% PEG8000 (native crystals) or 20% PEG3350 (selenomethionyl derivative crystals) as precipitant at pH 6.5. The native crystals belong to the space group *P*1 with cell parameters $a = 51.1 \text{ \AA}$, $b = 59.8 \text{ \AA}$, $c = 67.2 \text{ \AA}$, $\alpha = 96.6^\circ$, $\beta = 96.8^\circ$, $\gamma = 109.5^\circ$ and four molecules per asymmetric unit. The selenomethionyl derivative crystals belong to the space group *P*6₄22 with cell parameters $a = 73.8 \text{ \AA}$, $b = 73.8 \text{ \AA}$, $c = 123.4 \text{ \AA}$ with one molecule in the asymmetric unit.

X-ray diffraction data were collected from a single selenomethionyl derivative PA_N crystal at three wavelengths around the selenium K-edge at the Advanced Photon Source (APS), Chicago, and processed using HKL2000 (ref. 31) to 2.9 Å resolution. An additional high resolution native data set was collected in-house to 2.2 Å resolution using a Rigaku FR-E Cu-Kα rotating anode generator and R-Axis IV++ detector. The positions of six selenium atoms in the asymmetric unit were determined from the selenomethionyl derivative data sets using SHELX³². Initial phase estimates were obtained by MAD phasing with the program MLPHARE and improved by solvent flattening using DM at a nominal solvent content of 41%³³. The initial electron density map showed clear solvent boundaries, and it was possible to trace 80 residues of the four major helices. Following limited positional refinement using CNS 34, the working *R*-factor for this model for all amplitudes to 2.8 Å resolution had converged to 49%. After more than 50 cycles of manual main chain building and phase combination, the working *R*-factor for this model had decreased to 37%. This initial model was used for molecular replacement using PHASER³⁴ to obtain four clear solutions in the asymmetric unit of the native data set. Further refinement and rebuilding of this model was performed with the guidance of $2F_o - F_c$ and $F_o - F_c$ density maps using CNS³⁵. In the final stages, the refinement had converged to yield an *R*-factor of 23.1% (*R*_{free} of 25.2%) for all data in the resolution range 30.0–2.2 Å, with refinement of individual isotropic temperature factors and use of bulk solvent correction. Finally, water molecules were added by inspection of sA-weighted ($F_o - F_c$) and $2F_o - F_c$ electron density maps. A total of 107 water molecules were added to the model, and refinement and adjustment of these was performed until no further difference density could reasonably be ascribed to water. Inspection of difference Fourier maps in the later stages of refinement showed electron density (at $>4\sigma$) near D108, and the bound Mg²⁺ ion with three coordinated solvent molecules were carefully added into the model. When analysed using PROCHECK, the final structure has 89% of residues in the most favoured regions of the Ramachandran plot, and only seven residues are in disallowed regions as the result of reasonable hydrogen bonds.

Protease activity assay. The protease activity of PA_N was measured using a previously described protocol¹². In brief, Suc-LLVY-MCA (Sigma-Aldrich) was used as a substrate, and the fluorescence of 7-amino-4-methylcoumarin from peptide hydrolysis was measured by a Hitachi F-4500 fluorescence spectrophotometer at wavelength 405 nm. Chymotrypsin (Sigma-Aldrich) was used as a positive control.

31. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography* Vol. 276 (eds Carter, C. W. Jr & Sweet, R. M.) Part A 307–326 (Academic, 1997).

32. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).

33. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

34. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57**, 1373–1382 (2001).

35. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

LETTERS

The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit

Alexandre Dias^{1*}, Denis Bouvier^{1*}, Thibaut Crépin^{1*}, Andrew A. McCarthy^{1,2}, Darren J. Hart^{1,2}, Florence Baudin¹, Stephen Cusack^{1,2} & Rob W. H. Ruigrok¹

The influenza virus polymerase, a heterotrimer composed of three subunits, PA, PB1 and PB2, is responsible for replication and transcription of the eight separate segments of the viral RNA genome in the nuclei of infected cells. The polymerase synthesizes viral messenger RNAs using short capped primers derived from cellular transcripts by a unique 'cap-snatching' mechanism¹. The PB2 subunit binds the 5' cap of host pre-mRNAs^{2–4}, which are subsequently cleaved after 10–13 nucleotides by the viral endonuclease, hitherto thought to reside in the PB2 (ref. 5) or PB1 (ref. 2) subunits. Here we describe biochemical and structural studies showing that the amino-terminal 209 residues of the PA subunit contain the endonuclease active site. We show that this domain has intrinsic RNA and DNA endonuclease activity that is strongly activated by manganese ions, matching observations reported for the endonuclease activity of the intact trimeric polymerase^{6,7}. Furthermore, this activity is inhibited by 2,4-dioxo-4-phenylbutanoic acid, a known inhibitor of the influenza endonuclease⁸. The crystal structure of the domain reveals a structural core closely resembling resolvases and type II restriction endonucleases. The active site comprises a histidine and a cluster of three acidic residues, conserved in all influenza viruses, which bind two manganese ions in a configuration similar to other two-metal-dependent endonucleases. Two active site residues have previously been shown to specifically eliminate the polymerase endonuclease activity when mutated⁹. These results will facilitate the optimisation of endonuclease inhibitors^{10–12} as potential new anti-influenza drugs.

The exact role of the influenza polymerase PA subunit remains unclear, although it has been implicated in cap-binding, endonuclease activity, viral RNA binding and replication^{9,13–15}. PA is separable by trypsinization⁹ into a large carboxy-terminal domain—the crystal structure of which has recently been reported^{16,17}—and a small N-terminal domain, which contains residues important for protein stability, promoter binding, cap-binding and endonuclease activity of the polymerase complex⁹. In particular, the importance of Asp 108 and Lys 134 for endonuclease function called into question a previous conclusion that three acidic residues of PB1 were exclusively responsible for the endonuclease activity². The enzymology of the endonuclease within the context of intact viral ribonucleoprotein particles (RNPs) has been extensively studied. The endonuclease is activated by the binding of divalent cations, most effectively manganese followed by cobalt, and less efficiently magnesium, zinc and nickel, but not cadmium, calcium or trivalent ions⁶. RNPs bind four manganese ions in a highly cooperative fashion, presumably two each in the nucleotidyl-transferase and endonuclease sites⁶. It was concluded that the enzyme probably has a two-metal dependent mechanism in common with several other nucleases. Furthermore,

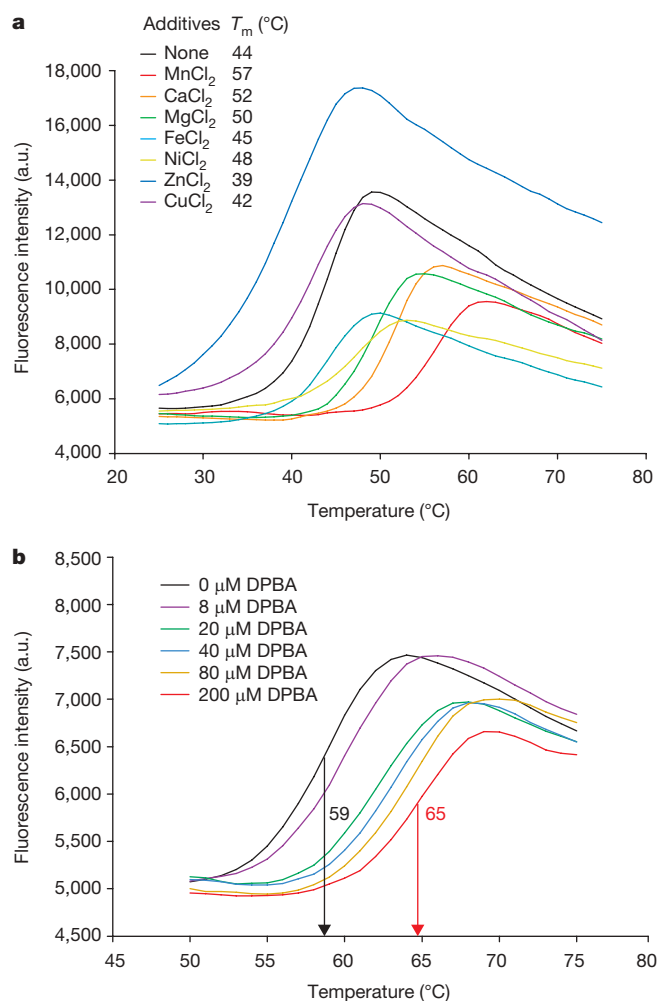


Figure 1 | Manganese ions stabilize PA-Nter. **a**, Assay of thermal stability using ThermoFluor. Thermal shift assays were performed with different metal salts. The results obtained in the absence of metal salts are indicated with a black line and those in the presence of metal salts with coloured lines. The different melting points (T_m) extracted from the curves at pH 8.0 are given in the inset. The effect of CoCl₂ on protein stability at pH 7.0 was investigated but was not interpretable owing to quenching by the metal. **b**, Assay of thermal stability with DPBA. ThermoFluor assays of PA-Nter were performed in the presence of MnCl₂ with increasing concentrations of DPBA. a.u., arbitrary units.

¹Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMR 5233, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France. ²Grenoble Outstation, European Molecular Biology Laboratory, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France.

*These authors contributed equally to this work.

the endonuclease cleaves single-stranded (ss)DNA with only slightly reduced activity compared to RNA; indeed capped ssDNA oligomers are substrates for cap-snatching⁷.

To clarify the functional role of the N-terminal domain of PA we studied a proteolytically stable fragment from residues 1 to 209 (denoted PA-Nter). To optimise the stability and mono-dispersity of PA-Nter we used dynamic light scattering and thermal shift assays with a variety of buffers and additives. These experiments showed a marked increase in thermal stability (apparent melting temperature, T_m , shifts from 44 to 57 °C) and improved mono-dispersity after the addition of manganese ions and to a lesser extent calcium and magnesium ions, but not other divalent metal ions (Fig. 1a). The structural effect of manganese binding, investigated by circular dichroism, demonstrated a small, but reproducible increase in helical content after the addition of 1 mM $MnCl_2$ (Supplementary Fig. 1). Because PA-Nter has been implicated in the viral endonuclease activity⁹, which is known to depend on divalent cations, notably manganese⁶, we assayed the domain for endonuclease activity using both structured

and unstructured ssRNA (Fig. 2) and ssDNA substrates (Supplementary Fig. 2). Using a partially structured 81-nucleotide panhandle RNA (ph-RNA) we found that PA-Nter has intrinsic RNase activity that is strictly divalent-cation-dependent (Fig. 2a, b). Consistent with previous results on RNPs⁶, strong activity was observed at pH 8 with manganese, weaker activity with magnesium and none with zinc, calcium or cobalt, whereas at pH 7 activity was also observed with cobalt (Fig. 2b). After 40 min incubation highly structured RNAs such as transfer RNA and signal recognition particle (SRP) *Alu* RNA are relatively resistant to degradation, partially structured ph-RNAs and short ph-RNAs are partially degraded and unstructured U-rich RNA is completely degraded, suggesting that the enzyme is single-strand specific (Fig. 2c). The enzyme also completely degrades circular ssDNA showing that it is a nonspecific endonuclease (Supplementary Fig. 2a). The endonuclease activity on both RNA and DNA is inhibited in a dose-dependent manner by the compound 2,4-dioxo-4-phenylbutanoic acid (DPBA), a known inhibitor of influenza endonuclease⁸ (Fig. 2d and Supplementary Fig. 2b). The inhibition

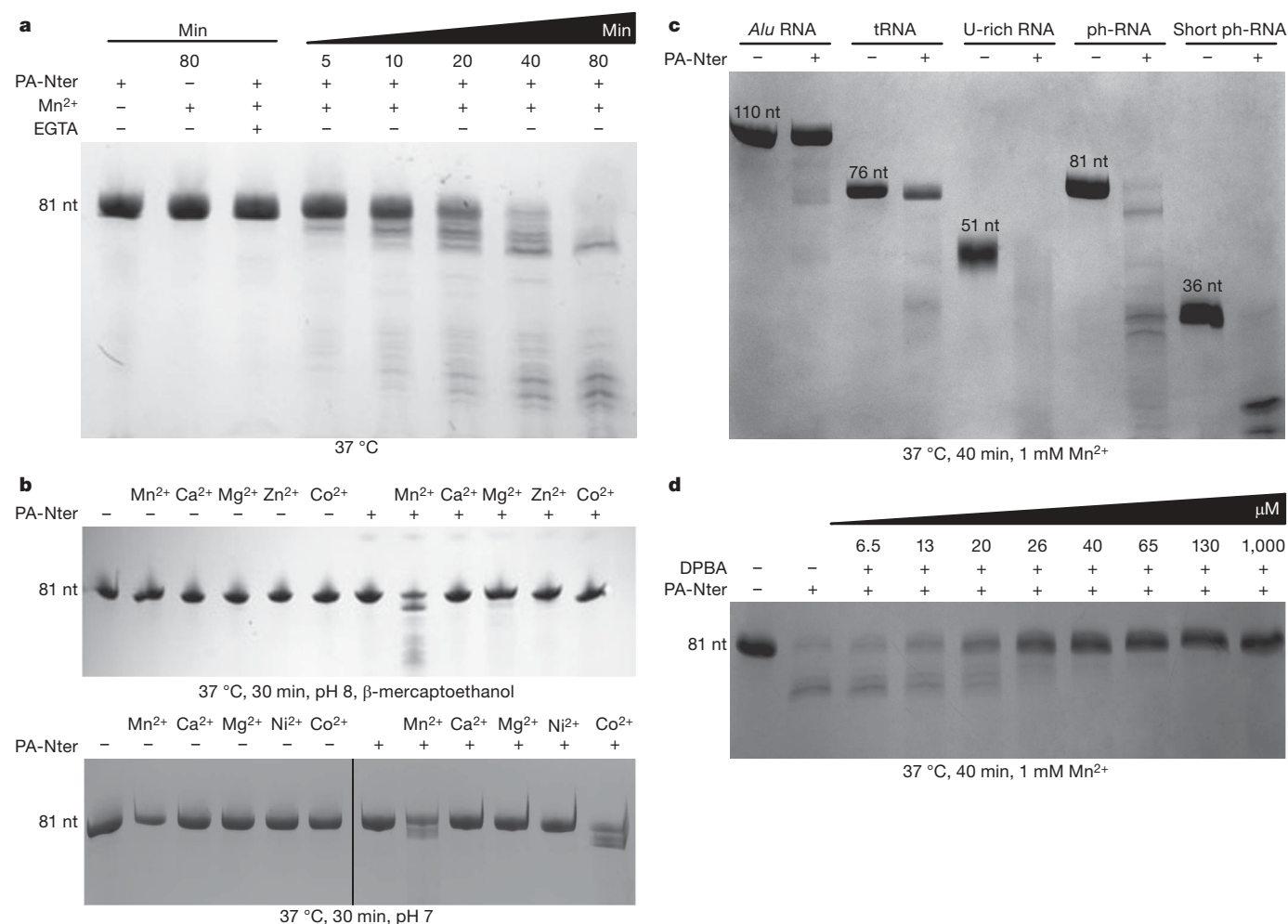


Figure 2 | Endonuclease activity of PA-Nter. **a**, A time series is shown. Purified ph-RNA was incubated with PA-Nter plus 1 mM $MnCl_2$. The incubation at 37 °C was stopped by adding 20 mM EGTA after 5, 10, 20, 40 and 80 min (lanes 4 to 8). As controls, ph-RNA was incubated for 80 min at 37 °C with PA-Nter only (lane 1), 1 mM $MnCl_2$ only (lane 2), or PA-Nter plus 1 mM $MnCl_2$ and 20 mM EGTA (lane 3). The reaction products were loaded on an 8% acrylamide and 8 M urea gel (15% for **b**, bottom panel and **c**) and stained with methylene blue. nt, nucleotides. **b**, The effect of divalent cations on PA-Nter RNase activity. In the top panel, purified ph-RNA plus PA-Nter were incubated at pH 8 in the presence of β -mercaptoethanol and 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $ZnCl_2$ or $CoCl_2$. In the bottom panel, ph-RNA and PA-Nter were incubated at pH 7 with 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $NiCl_2$ or $CoCl_2$. After 30 min the reactions were stopped by the addition of 20 mM

EGTA. Controls were performed using either salts or PA-Nter alone as indicated. At pH 7, $CoCl_2$ stimulated the endonuclease stronger than $MnCl_2$, whereas at pH 8, $CoCl_2$ precipitates and thus does not activate the endonuclease⁶. **c**, PA-Nter RNase activity on different RNA substrates. SRP *Alu* RNA, tRNA, U-rich RNA, ph-RNA or short ph-RNA were incubated with 1 mM $MnCl_2$ plus PA-Nter (lanes 2, 4, 6, 8 and 10) or in the absence of PA-Nter (lanes 1, 3, 5, 7 and 9). The digestion was performed at 37 °C. After 40 min the reaction was stopped by the addition of 20 mM EGTA. **d**, The inhibition of PA-Nter endonuclease activity by DPBA. Cleavage of ph-RNA by PA-Nter was tested at 37 °C during 40 min in the presence of 1 mM $MnCl_2$ and increasing concentrations of DPBA (0, 6.5, 13, 20, 26, 40, 65, 130 and 1,000 μM). As a control, ph-RNA was incubated with 1 mM $MnCl_2$ alone (lane 1).

constant (K_i) for this compound is estimated at 26 μ M, in excellent agreement with the half-maximal inhibitory concentration (IC_{50}) reported for the same compound inhibiting cleavage of capped RNA by the intact influenza virus polymerase⁸. Titrating the inhibitor to manganese-bound PA-Nter increases the thermal stability even further (apparent T_m shifts from 59 to 65 °C) (Fig. 1b), whereas the inhibitor has no effect on the stability of metal-free enzyme (data not shown).

We grew small square-plate crystals of PA-Nter only in the presence of both manganese and magnesium that diffracted to about 2 Å resolution, with three independent molecules in the asymmetric unit. The crystal structure reveals a single, folded domain with residues 1–196 visible, comprising seven α -helices and a mixed, five-stranded β -sheet (Fig. 3a). The structure-based sequence alignment among influenza A, B and C viruses (Fig. 3b) projected onto a surface

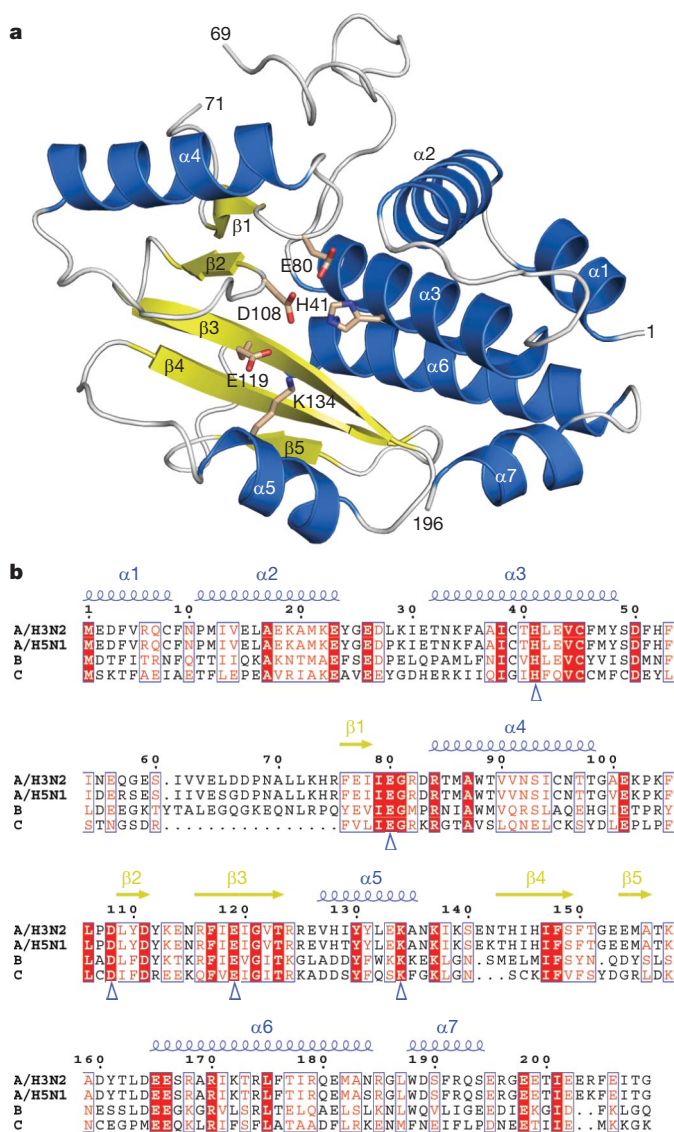


Figure 3 | Three-dimensional structure of PA-Nter. **a**, Ribbon diagram of the structure of influenza A/Victoria/3/1975 PA-Nter with α -helices in blue and β -strands in yellow, labelled as in **b**. The key active site residues are indicated in stick representation. **b**, Sequence alignment of PA-Nter from representative influenza strains: A/Victoria/3/1975 (human H3N2), A/Duck/Vietnam/1/2007 (avian H5N1), B/Ann Arbor/1/1966 (B) and C/Johannesburg/1/1966 (C). The secondary structure of PA-Nter of A/Victoria/3/1975 is shown over the sequence alignment in accordance with colours used for **a**. Residues in a solid red background are identical between the four sequences. The blue triangles indicate the key active site residues.

representation shows a very highly conserved depression that is strongly negatively charged owing to a concentration of acidic residues (Supplementary Fig. 3a, b), suggestive of an active site. A structure similarity search gave no high scoring hits, indicating that the global fold is new. The most similar protein found is the archaeal Holliday junction resolvase Hjc from *Pyrococcus furiosus*¹⁸. The structural alignment of PA-Nter with Hjc superposes helix α 3 and strands β 1– β 5 (Fig. 4a, b) encompassing a structural motif characteristic of many nucleases including resolvases and type II restriction enzymes. The motif includes catalytically important divalent metal ion binding acidic residues Asp 33 and Glu 46 of Hjc¹⁸, on which Asp 108 and Glu 119 of PA-Nter exactly superpose. Structural alignment of PA-Nter with type II restriction endonucleases such as BamHI or EcoRV shows a similar superposition of active site elements (Fig. 4c). The catalytically important residues Glu 45, Asp 74, Asp 90 and Lys 92 of EcoRV align with His 41, Asp 108, Glu 119 and Lys 134 of PA-Nter, respectively, although the lysines are positioned differently in the primary sequence (Fig. 4d)¹⁹. The conserved lysine is implicated in stabilizing the attacking hydroxide nucleophile during catalysis¹⁹. Thus PA-Nter is a new member of the PD-(D/E)XK nuclease superfamily which encompasses a diversity of enzymes involved in various aspects of DNA metabolism²⁰. In PA-Nter, the characteristic motif occurs at 107-PDLYDYK, although the separation between the two acidic residues is unusually short and the putative catalytically important lysine (Lys 134) has 'migrated' to an alternative position, as in some other members of the superfamily²⁰. In this family, PA-Nter is unusual in that it is biologically functional as an RNase and has a histidine in the active site.

To confirm that the conserved acidic residues of PA-Nter are metal binding residues we calculated an anomalous difference map using data collected at the manganese K absorption edge. Two manganese ions were identified in each active site as adjacent anomalous peaks separated by about 3.8 Å (Fig. 4e). The stronger peak (Mn1) is coordinated by Glu 80, Asp 108 and two water molecules; the weaker site (Mn2) by His 41, Asp 108, Glu 119 and the carbonyl oxygen of Ile 120. The cited residues are absolutely conserved in all influenza virus PA sequences (except for Ile 120, which is conservatively substituted) (Fig. 3b). The two metal sites correspond closely with those observed in restriction enzymes such as EcoRV (Fig. 4f). His 41 (positioned as Glu 45 in EcoRV) from helix α 3 could be important in conferring manganese specificity, because magnesium and calcium bind less readily to histidine. Manganese binding by His 41 and the resulting stabilization of helix α 3 could account for the extra helical content (estimated as 8–9 residues) detected after incubating PA-Nter with manganese (Supplementary Fig. 1). In the crystal, Mn1 and Mn2 are also coordinated by Glu 59 from a loop of an adjacent molecule. Superposition of DNA complexes of BamHI or EcoRV on PA-Nter shows that the Glu 59 carboxylate group corresponds closely to the position of the scissile phosphate group (Supplementary Fig. 4). Thus our structure mimics, to some extent, a substrate or product complex. A more precise understanding of the reaction mechanism and the roles of additionally conserved residues near the active centre (for example, Arg 84, Tyr 130 and Lys 137) must await structural studies with substrate or transition state analogues.

Our structural and biochemical results combined with previous observations on the trimeric polymerase provide compelling evidence that PA-Nter is the endonuclease that cleaves host mRNAs during cap-snatching. First, the domain has intrinsic RNA and DNA endonuclease activity that is preferentially activated by manganese, in accordance with observations reported for the viral RNPs (Fig. 2b)^{6,7}. Second, this activity is inhibited by a compound known to inhibit influenza endonuclease activity with a nearly identical K_i (Fig. 2d)⁸. Third, the domain contains a structural motif characteristic of the catalytic core of a broad family of nucleases, including type II endonucleases. The active site features a cluster of three acidic residues (Glu 80, Asp 108 and Glu 119) and a putative catalytic lysine (Lys 134) (Fig. 4). Fourth, these acidic residues, together with His 41,

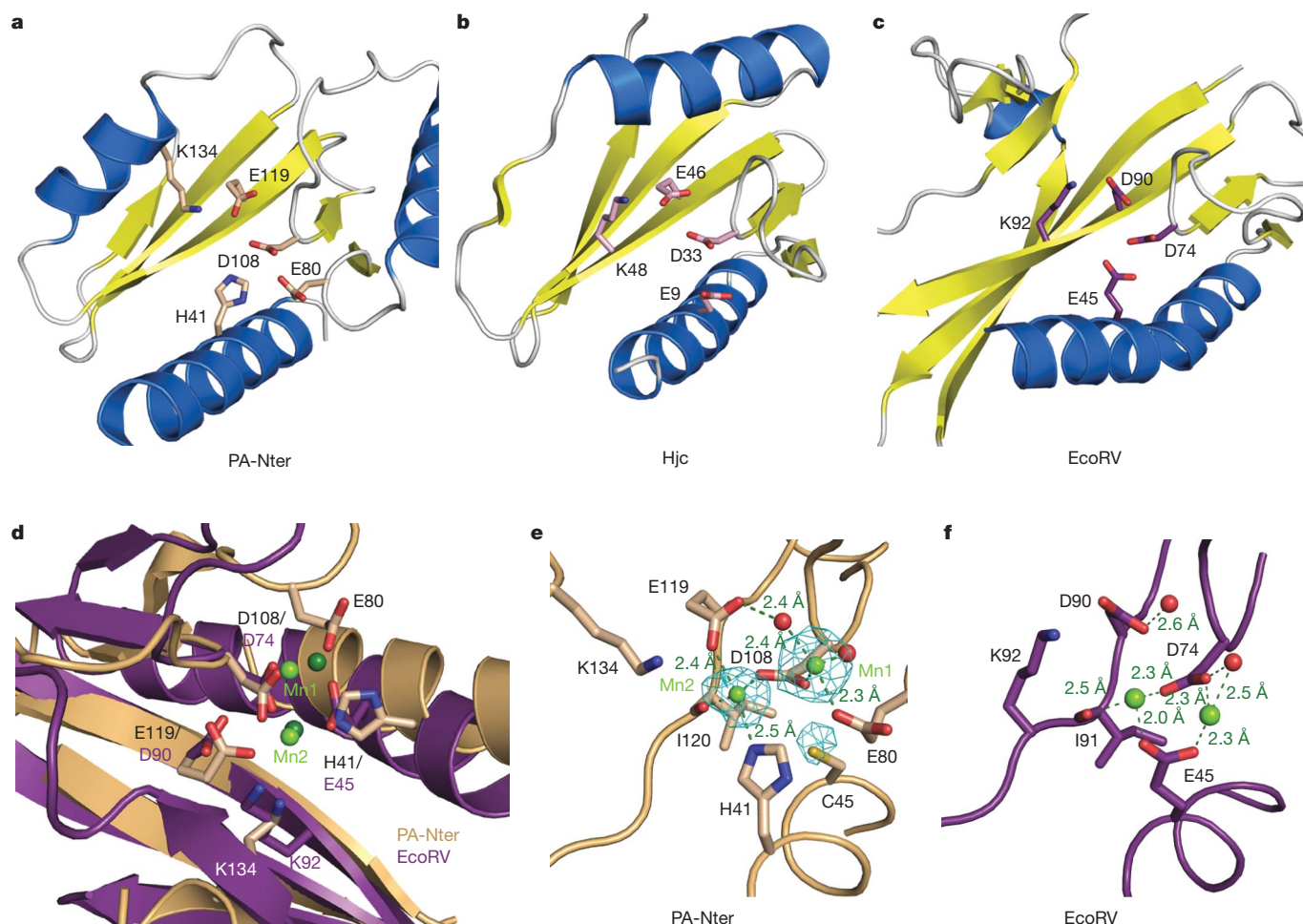


Figure 4 | Comparison of PA-Nter with other nucleases of the PD-(D/E)XK superfamily. **a–c**, Comparison of PA-Nter (**a**), *P. furiosus* Holliday junction resolvase Hjc (PDB accession 1GEF)¹⁸ (**b**), and *E. coli* EcoRV restriction enzyme (PDB accession 1STX, product complex with DNA and manganese)¹⁹ (**c**) after superposition of the conserved core active site structural motif. The root mean squared deviations are 2.9 Å for 77 aligned C α atoms of Hjc (Z-score 3.3), and 2.46 (3.1) Å for 55 (72) aligned C α atoms of EcoRV. Secondary-structure elements are as in Fig. 3a with key active site residues in stick representation. **d–f**, Details of the metal-binding active sites of influenza PA-Nter (molecule A) (**d**) and *E. coli* EcoRV restriction

enzyme product complex (**e**, **f**). **d**, Superposition of the active sites showing overlap of key active site metal binding and catalytic functional groups of the two proteins. Manganese ions are light green for PA-Nter and dark green for EcoRV. **e**, **f**, Manganese ion interactions within the active sites of PA-Nter (**e**) and the EcoRV product complex (**f**). Manganese ions and water molecules are shown as green and red spheres, respectively. In **e** the anomalous difference map contoured at 3 σ , calculated using manganese K edge diffraction data and model phases, is in cyan. Peak heights are 14.1, 10.1 and 5.0 σ for Mn1, Mn2 and the sulphur of Cys 45, respectively.

all absolutely conserved in influenza viruses, coordinate two manganese ions in a configuration consistent with a two-metal-dependent reaction mechanism as proposed for many nucleases (Fig. 4e)^{19,21,22}. Fifth, mutations of important active site residues D108A and K134A have been shown to specifically abolish endonuclease activity of the trimeric polymerase⁹. In addition, the mutation H41A abolished all polymerase activity, consistent with a crucial role in both the stability of PA and the activity of the endonuclease⁹.

Our results establish for the first time, to our knowledge, a unique role for the PA subunit of influenza virus polymerase and contradict the widely held view that the endonuclease active site is located in the PB1 subunit. This was on the basis of data showing that PB1 residues 508–522 could be cross-linked to capped RNAs by means of a thioluridine adjacent to the endonucleolytic cleavage site². The cross-linked region includes acidic residues Glu 508, Glu 519 and Asp 522, which when individually mutated to alanine specifically abolished endonuclease activity, although their exact function was not determined. These data can be reconciled with our observations if the cross-linked region of PB1 is in spatial proximity to the actual endonuclease active site in PA-Nter. We note that similar cross-linking experiments² led to the incorrect identification of the cap-binding site on PB2 (refs 4, 23).

Indeed it is very likely that other regions of the polymerase regulate the intrinsic endonuclease activity of PA-Nter, because it is known that there is a cooperative functional interaction between capped RNA binding, 3' and 5' end viral RNA (vRNA) binding, and endonuclease activity^{24,25}. It is therefore surprising that, as with the cap-binding domain of PB2, a small independently folded domain should have so many of the functional properties of the endonuclease reported for the trimeric complex, as both cap-snatching activities were thought to be detectable only in the trimeric complex. Nonetheless, full understanding of the concerted mechanism of cap-dependent transcription will undoubtedly require high resolution structures of the intact polymerase with bound vRNA and capped RNA.

Influenza can be a serious illness and there is a continuing risk of a devastating pandemic resulting from the eventual acquisition by highly pathogenic avian strains (for example, H5N1) of the ability to be transmissible from human to human. Whereas mutations in the viral polymerase are known to have an important role in virulence and interspecies transmission^{26,27}, none of the systematic differences in PA-Nter between human and avian strains (Fig. 3b) has been implicated in these processes nor seems to be critically placed as to affect endonuclease function. Several specific influenza virus endonuclease inhibitors

have been described^{18,10,11} including some designed to chelate metal ions bound in the active site¹². Our results show that at least one of these inhibitors is active against PA-Nter and that it binds and stabilizes the metal-bound form of the enzyme (Fig. 1b). These observations will be helpful in developing potential new antivirals using a structure-based approach.

METHODS SUMMARY

PA-Nter, residues 1–209 from A/Victoria/3/1975 (H3N2), was expressed in *Escherichia coli* and purified by affinity and gel filtration chromatography. The influence of metal ions on thermal stability was tested by ThermoFluor assays²⁸. The endonuclease activity was tested by incubation at 37 °C of 13 µM PA-Nter with 10 µM of various RNA substrates: *Alu* RNA, 110 nucleotides of the *Alu* domain of *Pyrococcus horikoshii* SRP RNA, *Candida albicans* tRNA^{Asn}, U-rich RNA (5'-GGCCAUCUGU₇ CCCU₁₁CU₁₉-3')²⁹, ph-RNA of 81 nucleotides³⁰, and short ph-RNA of 36 nucleotides comprising just the conserved 3' and 5' ends with a short linker and circular single stranded DNA (M13mp18) (Fermentas). Crystals diffracting to a maximum of 2 Å resolution were obtained at 20 °C by the hanging-drop method using a protein solution of 5–10 mg ml⁻¹ in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 2.5 mM MnCl₂ and a reservoir composition of 1.2 M Li₂SO₄, 100 mM MES, pH 6.0, 10 mM Mg acetate and 3% ethylene glycol. Diffraction data were collected on beamlines ID14-4 and ID23-1 at the European Synchrotron Radiation Facility (ESRF). The structure was solved by the single-wavelength anomalous diffraction (SAD) method using a gadolinium-chloride-soaked crystal, and initial phases were improved by three-fold non-crystallographic symmetry averaging. Data were also measured on a native crystal at the manganese K edge to reveal the location and identity of bound manganese ions through anomalous difference Fourier synthesis. The structure described is that of a native crystal soaked with additional MnCl₂ at 10 mM. The final *R* factor (*R*_{free}) is 0.217 (0.268). Of the three molecules in the asymmetric unit, A has the best defined metal structure and D is the most disordered. Further details are available in the Supplementary Methods including a table of crystallographic statistics (Supplementary Table 1).

Received 10 November; accepted 22 December 2008.

Published online 1 February 2009.

- Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847–858 (1981).
- Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20**, 2078–2086 (2001).
- Fechter, P. *et al.* Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J. Biol. Chem.* **278**, 20381–20388 (2003).
- Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15**, 500–506 (2008).
- Shi, L., Summers, D. F., Peng, Q. & Galarz, J. M. Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. *Virology* **208**, 38–47 (1995).
- Doan, L., Handa, B., Roberts, N. A. & Klumpp, K. Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. *Biochemistry* **38**, 5612–5619 (1999).
- Klumpp, K., Doan, L., Roberts, N. A. & Handa, B. RNA and DNA hydrolysis are catalyzed by the influenza virus endonuclease. *J. Biol. Chem.* **275**, 6181–6188 (2000).
- Tomassini, J. *et al.* Inhibition of cap (m⁷GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob. Agents Chemother.* **38**, 2827–2837 (1994).
- Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80**, 7789–7798 (2006).
- Hastings, J. C., Selnick, H., Wolanski, B. & Tomassini, J. E. Anti-influenza virus activities of 4-substituted 2,4-dioxobutanoic acid inhibitors. *Antimicrob. Agents Chemother.* **40**, 1304–1307 (1996).
- Tomassini, J. E. *et al.* A novel antiviral agent which inhibits the endonuclease of influenza viruses. *Antimicrob. Agents Chemother.* **40**, 1189–1193 (1996).
- Parkes, K. E. *et al.* Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. *J. Med. Chem.* **46**, 1153–1164 (2003).
- Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76**, 8989–9001 (2002).
- Lee, M. T. *et al.* Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic Acids Res.* **30**, 429–438 (2002).
- Honda, A., Mizumoto, K. & Ishihama, A. Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl Acad. Sci. USA* **99**, 13166–13171 (2002).
- He, X. *et al.* Crystal structure of the polymerase PA_C–PB1_N complex from an avian influenza H5N1 virus. *Nature* **454**, 1123–1126 (2008).
- Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454**, 1127–1131 (2008).
- Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. & Morikawa, K. Crystal structure of the archaeal Holliday junction resolvase Hjc and implications for DNA recognition. *Structure* **9**, 197–204 (2001).
- Horton, N. C. & Perona, J. J. DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **43**, 6841–6857 (2004).
- Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L. & Ginalski, K. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* **7**, 40 (2007).
- Beese, L. S. & Steitz, T. A. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33 (1991).
- Viadiu, H. & Aggarwal, A. K. The role of metals in catalysis by the restriction endonuclease BamHI. *Nature Struct. Biol.* **5**, 910–916 (1998).
- Tarendeau, F. *et al.* Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* **4**, e1000136 (2008).
- Hagen, M., Chung, T. D., Butcher, J. A. & Krystal, M. Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J. Virol.* **68**, 1509–1515 (1994).
- Rao, P., Yuan, W. & Krug, R. M. Crucial role of CA cleavage sites in the cap-snatching mechanism for initiating viral mRNA synthesis. *EMBO J.* **22**, 1188–1198 (2003).
- Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl Acad. Sci. USA* **102**, 18590–18595 (2005).
- Salomon, R. *et al.* The polymerase complex genes contribute to the high virulence of the human H5N1 influenza virus isolate A/Vietnam/1203/04. *J. Exp. Med.* **203**, 689–697 (2006).
- Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. ThermoFluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–298 (2006).
- Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. & Gale, M. Jr. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454**, 523–527 (2008).
- Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13**, 3158–3165 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the ESRF, EMBL and MRC-France for access to synchrotron facilities and the Partnership for Structural Biology for an integrated structural biology environment. The work was partially funded by the EU FLUPOL contract (SP5B-CT-2007-044263) and the ANR FLU INTERPOL contract (ANR-06-MIME-014-02). A.D. has a PhD fellowship from the MENRT. We thank J.-L. Decout for advice about the inhibitor, E. Kowalinski and C. Swale for RNA transcripts, and C. Petosa and J. Perona for a critical reading of the manuscript.

Author Contributions A.D., D.B., T.C. and A.A.M. performed the experiments, D.J.H. and F.B. supervised the search for a soluble fragment of PA, R.W.H.R. and S.C. supervised the biochemical characterization of PA-Nter, and S.C. supervised the crystallography. S.C. wrote the paper with the help of A.D., D.B., T.C. and R.W.H.R.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank (PDB) under accession codes 2W69 and R2W69SF. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C. (cusack@embl.fr).

An unusual mechanism of thymidylate biosynthesis in organisms containing the *thyX* gene

Eric M. Koehn¹, Todd Fleischmann¹, John A. Conrad², Bruce A. Palfrey², Scott A. Lesley³, Irimpan I. Mathews⁴ & Amnon Kohen¹

Biosynthesis of the DNA base thymine depends on activity of the enzyme thymidylate synthase to catalyse the methylation of the uracil moiety of 2'-deoxyuridine-5'-monophosphate. All known thymidylate synthases rely on an active site residue of the enzyme to activate 2'-deoxyuridine-5'-monophosphate^{1,2}. This functionality has been demonstrated for classical thymidylate synthases, including human thymidylate synthase, and is instrumental in mechanism-based inhibition of these enzymes. Here we report an example of thymidylate biosynthesis that occurs without an enzymatic nucleophile. This unusual biosynthetic pathway occurs in organisms containing the *thyX* gene, which codes for a flavin-dependent thymidylate synthase (FDTs), and is present in several human pathogens³⁻⁵. Our findings indicate that the putative active site nucleophile is not required for FDTs catalysis, and no alternative nucleophilic residues capable of serving this function can be identified. Instead, our findings suggest that a hydride equivalent (that is, a proton and two electrons) is transferred from the reduced flavin cofactor directly to the uracil ring, followed by an isomerization of the intermediate to form the product, 2'-deoxythymidine-5'-monophosphate. These observations indicate a very different chemical cascade than that of classical thymidylate synthases or any other known biological methylation. The findings and chemical mechanism proposed here, together with available structural data, suggest that selective inhibition of FDTs, with little effect on human thymine biosynthesis, should be feasible. Because several human pathogens depend on FDTs for DNA biosynthesis, its unique mechanism makes it an attractive target for antibiotic drugs.

Classical thymidylate synthases, encoded by the *thyA* and *TYMS* genes, are present in most eukaryotes, including humans, and are frequently targeted by chemotherapeutic and antibiotic drugs. A recently discovered class of thymidylate synthases, the FDTs^{3,6,7}, is encoded by the *thyX* gene and has been found primarily in prokaryotes and viruses^{3,8}, including several pathogens and biological warfare agents (see <http://www.cdc.gov>). Several organisms, including human pathogens, rely solely on *thyX* for thymidylate synthesis (for example, all *Rickettsia* lack genes coding for DHFR, thymidylate synthase, and thymine kinase). It has recently been suggested that *thyX* limits chromosomal replication in these organisms⁹. FDTs share no structure or sequence homology with classical thymidylate synthases, and thus present a promising new frontier for antibacterial/antiviral drug development⁵⁻⁷.

The catalytic mechanism of classical thymidylate synthases is presented in Fig. 1a^{1,2}. A strictly conserved active site cysteine covalently activates the uracil ring by nucleophilic Michael addition at the C6 position of 2'-deoxyuridine-5'-monophosphate (dUMP, Fig. 1a, step 2). The resulting enolate then attacks (step 3) the iminium form of

N⁵,N¹⁰-methylene-5,6,7,8-tetrahydrofolate (CH₂H₄folate), followed by elimination of tetrahydrofolate (H₄folate, step 4). Finally, a hydride transfer from C6 of the H₄folate yields the products 2'-deoxythymidine-5'-monophosphate (dTMP) and dihydrofolate (H₂folate, step 5). Common drugs that target classical thymidylate synthase either covalently bind to the catalytic nucleophile (for example, 5F-dUMP) or non-covalently bind the folate binding pocket (for example, Tomudex).

It has previously been proposed that FDTs have a chemical mechanism analogous to the classical thymidylate synthase mechanism, but with a serine residue acting as a nucleophile (Fig. 1b, step 2) and a flavin cofactor providing a hydride to terminate the reaction (Fig. 1b, step 6). This results in the production of H₄folate rather than H₂folate^{10,11}. The suggestion that a serine serves as the active site nucleophile in FDTs was originally based on sequence alignments of *thyX* genes that indicated no conserved cysteine but a strictly conserved serine. Crystal structures of FDTs from three very different organisms^{7,12,13} placed this conserved serine about 4 Å from the C6 position of dUMP (for example, Fig. 2a). However, without a neighbouring general base this serine will not be deprotonated, decreasing its potential reactivity, casting doubt on this serine's putative role as a nucleophile.

Point mutation studies were performed with FDTs from *Helicobacter pylori* (HpFDTs), where the conserved serine residue was mutated to either alanine or cysteine (S84A and S84C). Although both mutations were found to retain activity⁴, it was assumed that an adjacent serine (Ser85) could have rescued the activity of S84A. Abolished enzyme activity of a double mutant (S84A/S85A) supported this hypothesis. Furthermore, matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry analysis showed that the S84C mutant forms a covalent adduct with dTMP. These results were used to propose that Ser 84 activates dUMP (Fig. 1b)^{4,11}.

To test this hypothesis further, we performed similar mutation studies using FDTs from *Thermotoga maritima* (TmFDTs). The active site of this enzyme contains a strictly conserved serine, Ser88, and no alternative nucleophilic residues (see refs 6, 7 and Supplementary Information). Activity tests for S88A and S88C (for details see ref. 11 and Supplementary Information) indicated that both mutants were still active. Possible contamination by classical *Escherichia coli* thymidylate synthase was ruled out by a series of control experiments to ensure that mutant FDTs was the sole source of the observed activity (see Supplementary Information). The FDTs activity of these mutants and the lack of classical thymidylate synthase activity of S88C demonstrate that the conserved active site serine does not serve as a catalytic nucleophile in the FDTs reaction, in contrast to the mechanism proposed hitherto (Fig. 1b).

¹Department of Chemistry, University of Iowa, Iowa City, Iowa 52242, USA. ²Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ³The Joint Center for Structural Genomics at the Genomics Institute of Novartis Research Foundation, San Diego, California 92121, USA. ⁴Stanford Synchrotron Radiation Laboratory, Stanford University, Menlo Park, California 94025, USA.

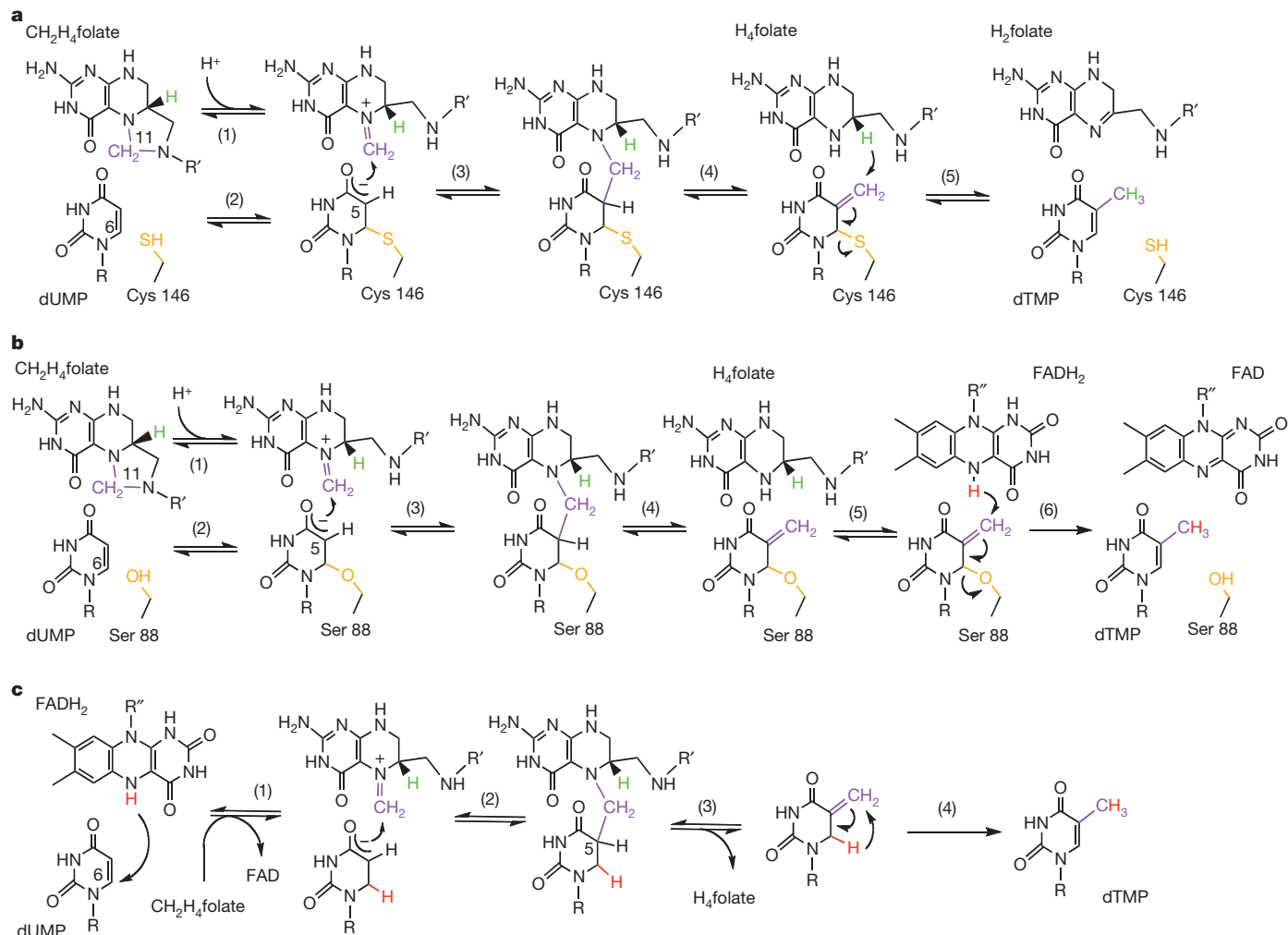


Figure 1 | Thymidylate synthase mechanisms. **a**, The chemical mechanism for the classical thymidylate synthase catalysed reaction^{1,2}. **b**, The chemical mechanism for the FDTs proposed hitherto¹¹. **c**, The newly proposed mechanism for the FDTs that does not rely on an enzymatic nucleophile.

Crystal structures of the S88A and S88C mutants were obtained at 1.95 and 2.05 Å resolution, respectively (see data collection/refinement statistics in Supplementary Information), and their electron density is compared to that of the wild-type *Tm*FDTs (Fig. 2). The electron densities for both mutants indicate minimal changes in folding and active site configuration. It is apparent from Fig. 2c that there is no covalent bond between the cysteine and dUMP in the crystal. Nevertheless, a MALDI-TOF analysis of the trypsin-digested S88C *Tm*FDTs indicated that Cys 88 is bound to dUMP (see Supplementary Information), as previously reported for *Hp*FDTs⁴. Even in solution, cysteine covalently binds to the C6 position of uracil¹⁴. These facts, together with the low activity of S88C, suggest that the observed Cys 88–dUMP complex is not part of the FDTs catalytic cascade, but rather an inhibitory dead-end complex.

A critical piece of evidence for the covalent bond between the active site cysteine in classical thymidylate synthase and dUMP is the crystal structure of a covalently bound 5-fluoro-dUMP (5F-dUMP) in complex with CH₂H₄folate (Protein Data Bank accession 1t1s, ref. 15). In contrast to classical thymidylate synthase, FDTs does not covalently bind 5F-dUMP upon incubation with CH₂H₄folate, as confirmed by both MALDI-TOF analysis (see Supplementary Information) and crystal structure analysis (Protein Data Bank 1o28, ref. 7) obtained under similar conditions. Another test for similar Michael addition to the C6 of dUMP is the dehalogenation of 5-bromo-dUMP (5Br-dUMP)^{2,16}. This test resulted in no reactivity

with FDTs (see Supplementary Information). We also solved the crystal structures of *Tm*FDTs with FAD and both 5-halogenated-dUMPs (Protein Data Bank 1o27 and 1o28); the 5Br-dUMP–FAD structure for FDTs from *Mycobacterium tuberculosis* (Protein Data Bank 2af6) was solved by ref. 13. These structures are nearly the same as the complex with dUMP (Fig. 2a) and do not support a nucleophilic attack of any enzyme residue on the C6 of dUMP. These observations emphasize the distinctions between the mechanisms of classical thymidylate synthase and FDTs, and in light of the activity of the S88A mutant, support a mechanism in which FDTs does not involve a Michael addition of an enzymatic nucleophile.

To expose the nature of the FDTs catalysed reaction we followed the flow of hydrogens along the catalytic pathway by isotopic substitution of a specific hydrogen. We have previously found that when conducting the FDTs reaction in D₂O (50% D), deuteration of the reduced flavin leads to deuterated dTMP (using electron spray ionization mass spectrometry, ESI-MS, analysis), and that reaction with tritiated 6T-CH₂H₄folate yields 6T-H₄folate¹¹. These results contrast the same experiments with classical thymidylate synthases, where reactions performed in D₂O do not incorporate deuterium into the dTMP and the labelled hydride from CH₂H₄folate always transfers to the dTMP¹⁷. In the past, we and others^{4,11} suggested that these findings support the mechanism illustrated in Fig. 1b, but the current findings contradict that mechanism and required further tests. By repeating the experiment in D₂O (this time >99.6% D), and analysing the product

with FDTs (see Supplementary Information). We also solved the crystal structures of *Tm*FDTs with FAD and both 5-halogenated-dUMPs (Protein Data Bank 1o27 and 1o28); the 5Br-dUMP–FAD structure for FDTs from *Mycobacterium tuberculosis* (Protein Data Bank 2af6) was solved by ref. 13. These structures are nearly the same as the complex with dUMP (Fig. 2a) and do not support a nucleophilic attack of any enzyme residue on the C6 of dUMP. These observations emphasize the distinctions between the mechanisms of classical thymidylate synthase and FDTs, and in light of the activity of the S88A mutant, support a mechanism in which FDTs does not involve a Michael addition of an enzymatic nucleophile.

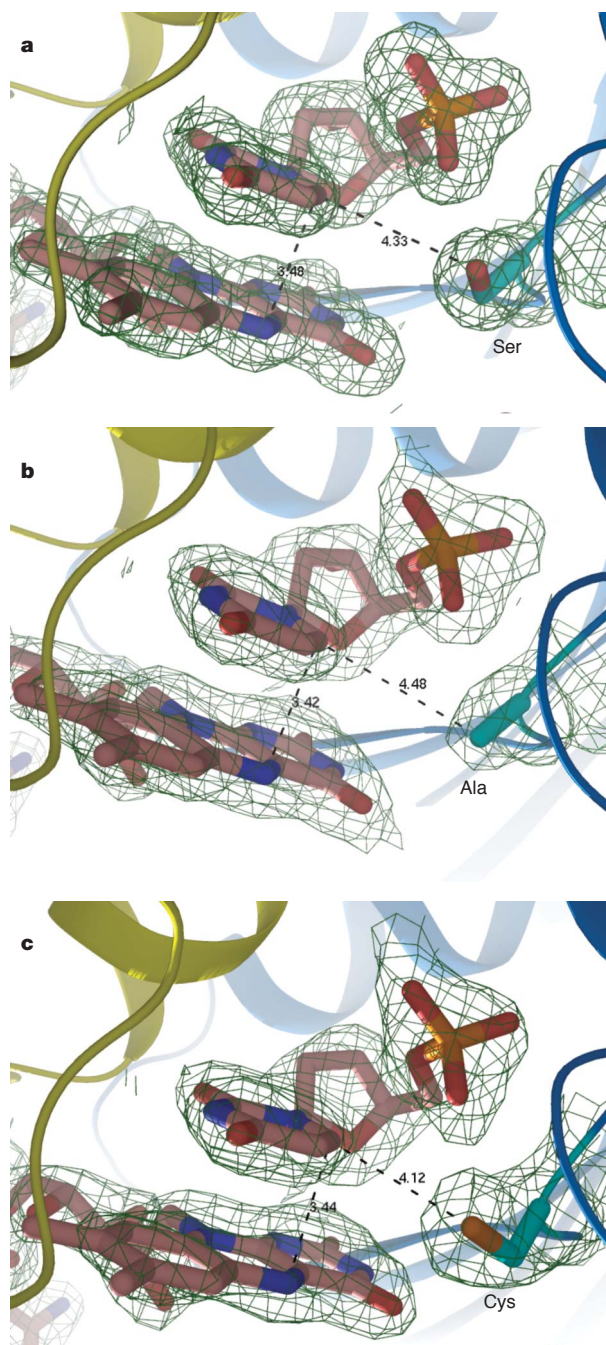


Figure 2 | Crystal structures of the FdTS-FAD-dUMP complex. **a**, Wild-type *TmFdTS* (Protein Data Bank 1o26); **b**, S88A mutant (Protein Data Bank 3g4a); **c**, S88C mutant (Protein Data Bank 3g4c). The distance between the C6 carbon of dUMP and the reducing centre of the flavin (N5 of FAD) is 3.4 Å for all three enzymes. The distances of the side chain of residue 88 to C6 are 4.3, 4.5 and 4.1 Å, for wild-type FdTS, S88A and S88C, respectively. The electron density maps are $2F_o - F_c$ with a contour level of 1.0 sigma.

using ESI-MS, ^1H NMR and ^2H NMR, we found that at 65 °C (close to the physiological temperature of *T. maritima*) the product was indeed deuterated at the C7 position (Fig. 3a, b). However, when we performed the same experiment at 37 °C, NMR analyses indicated the formation of both 6D-dTMP (60%) and 7D-dTMP (40%) (Fig. 3c, d). This result is quite intriguing, as no mechanism previously proposed for FdTS predicts formation of 6D-dTMP, and such a phenomenon has never been reported for any thymidylate synthase reaction.

The lack of an obvious enzymatic nucleophile and the ability to trap deuterium from D_2O at C6 of the product demonstrate that the

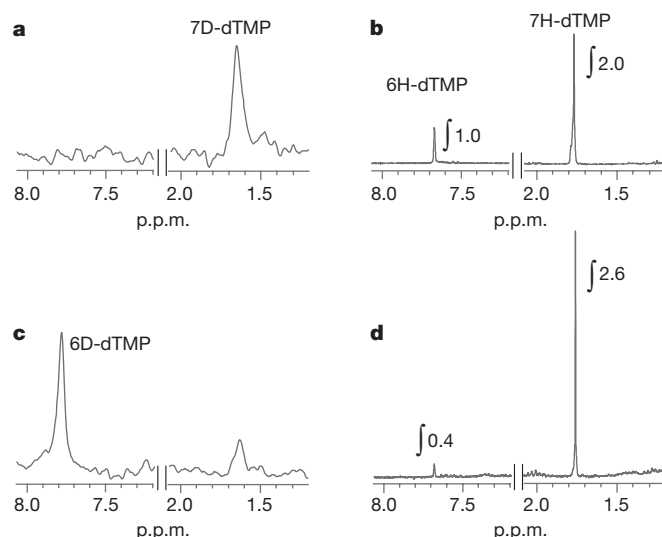


Figure 3 | ^2H NMR (a, c) and ^1H NMR (b, d) spectra of dTMP produced in the FdTS catalysed reaction of dUMP in D_2O (experiment A in Fig. 4). Spectra a and b were from the reaction at 65 °C; spectra c and d were from the same reaction at 37 °C. The latter spectra clearly indicate the presence of 6D-dTMP (~60%). p.p.m., parts per million. The integrals (∫) in b and d indicate the relative area under the peaks.

chemical mechanisms of FdTS and classical thymidylate synthase differ substantially. Without an enzymatic nucleophile, the FdTS catalysed reaction could proceed via Michael addition of a hydroxide ion or through participation of the flavin prosthetic group. For hydroxide to serve as a nucleophile, a water molecule must be activated by a general base in the active site (for example, the catalytic triad in hydrolytic enzymes). All crystal structures of FdTSs indicate that there is no such basic system available in the active site. Additional experiments using reduced 5-carba-5-deaza-FAD resulted in dTMP formation, excluding the possibility that the reduced N5 of FADH_2 is the nucleophile. Notably, the FdTS mechanism requires a hydrogen transfer to the C6 of the uracil moiety to explain the formation of 6D-dTMP from reactions performed in D_2O , which is inconsistent with either hydroxyl or flavin as Michael nucleophiles.

In Fig. 1c we propose a new chemical mechanism consistent with current data and previous findings¹⁸, wherein a hydride equivalent from the N5 of FADH_2 is transferred to C6 of dUMP (Fig. 1c, step 1). The resulting enolate anion nucleophilically attacks the iminium methylene of $\text{CH}_2\text{H}_4\text{folate}$, and an elimination of H5 from dUMP and H_4folate results in a $\text{C5}=\text{C7}$ double bond (steps 2 and 3). This exocyclic-methylene intermediate then isomerizes to form the product, dTMP (step 4). The intermediate proposed here is unique in nucleotide biochemistry, but this isomer of the thymine moiety is chemically feasible and quite stable in solution¹⁹. This mechanism is compatible with our previous findings¹⁸ on the oxidative half-reaction if the equilibrium constant for the first step lies to the left. Because we have no experimental data regarding the methylene transfer and the initial activation (if not H transfer), steps 2 and 3 are proposed here as a logical path towards the product and step 1 might be preceded by other activation steps.

Because the isomerization of the putative intermediate (Fig. 1c, step 4) does not occur rapidly in solution¹⁹, the enzyme could catalyse this transformation by the two mechanisms illustrated in Fig. 4. An enzymatic acid could catalyse this step via an addition-elimination mechanism (AEM), in which a proton is added to the $\text{C5}=\text{C7}$ double bond and the intermediate cation loses a proton from C6 to form the product. Alternatively, the thermodynamic driving force ($>6 \text{ kcal mol}^{-1}$ as estimated from semi-empirical quantum mechanical calculations) could favour a 1,3-sigmatropic rearrangement (1,3-hydride shift)²⁰.

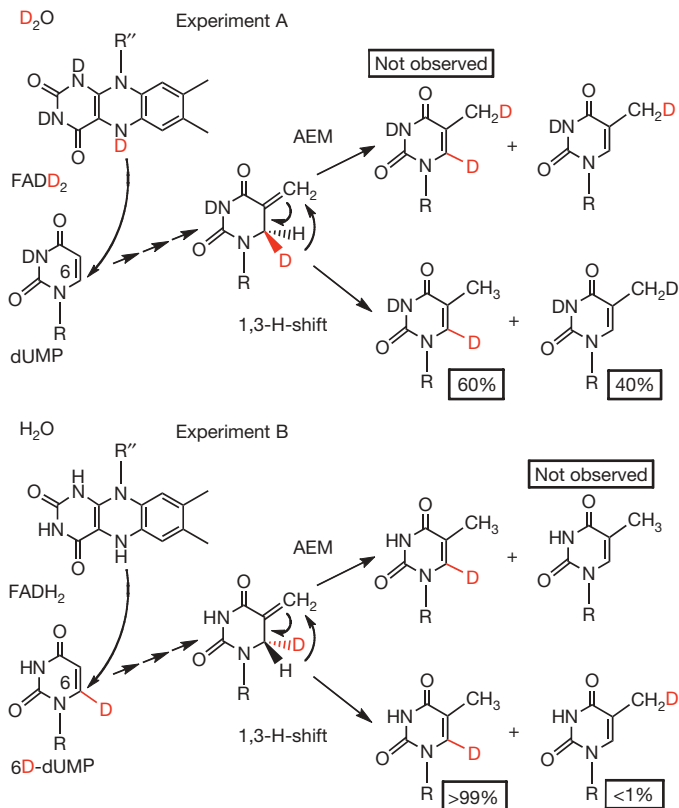


Figure 4 | Hydride flow. An illustration of two experimental approaches to examine the hydride flow in the reaction catalysed by the thermophilic *TmFDTs* at reduced temperature ($37^\circ C$). Experiment A was performed in a D_2O buffer using 6H-dUMP (that is, unlabelled dUMP). Experiment B was performed in an H_2O buffer using 6D-dUMP (see Supplementary Information). Percentages below each species represent the relative quantities of product formation as indicated by 1H and 2H NMR.

When conducting the reaction in D_2O , an AEM would lead to 6D,7D-dTMP, but ESI-MS analysis (Supplementary Fig. 4) did not indicate such a product. Therefore, a suitable explanation is an enzyme-catalysed isomerization via a 1,3-H shift (Fig. 4, lower path).

The finding that two different isotopically labelled products are formed at $37^\circ C$ was intriguing and warranted further investigation. Thus, we performed the FDTs reactions using dUMP, with D at its C6, in an H_2O buffer at $37^\circ C$ (Fig. 4, experiment B). NMR analysis of the product showed the formation of 6D-dTMP (>99%). If any 6H-dTMP was formed, it was below our detection limits (<1%). The observations for both experiments (6H-dUMP in D_2O and 6D-dUMP in H_2O) can be explained by the combination of normal kinetic isotope effect (KIE; H reacts faster than the heavier D) and reduced stereoselectivity at reduced temperature ($37^\circ C$). Lack of stereoselectivity at reduced temperature has already been observed during the reductive-half reaction of FDTs, which transfers both 4-(R) and 4-(S) hydride of NADPH¹¹. A KIE of 10 and stereoselectivity of 87%, for example, would result in production of 60:40 C6:C7-dTMP when 6H-dUMP reacts with FDTs in D_2O , and more than 99:1 C6:C7-dTMP produced when 6D-dUMP reacts in H_2O (see Supplementary Information for a more detailed discussion).

The proposed hydrogen transfer to the C6 of dUMP by the flavin cofactor is further supported by the crystal structures, which show a short distance (3.4 Å) between N5 of the flavin ring and C6 of dUMP (Fig. 2). Such hydride transfer from FADH₂ to a uracil ring is an atypical chemistry for thymidylate synthases and nucleotide methylation in general, but is not unprecedented in enzymology. For example, dihydroorotate dehydrogenase and old yellow enzyme are other flavo-proteins that catalyse similar chemistry^{21,22}.

To the best of our knowledge, neither hydride transfer to the uracil ring nor an isomerization of such an intermediate has been reported for any thymine biosynthetic pathway or other nucleotide methylations. Notably, such a chemical mechanism is very different from that of classical thymidylate synthases, and along with structural differences, may help explain why classical thymidylate synthase inhibitors have a reduced effect on FDTs⁵. These findings suggest that selective inhibition of FDTs should be feasible, and may further alleviate the constraint of an enzymatic nucleophile from structure-based rational drug design efforts. Rationally designed compounds could mimic the non-covalently bound intermediate or the transition states for its formation and isomerization. Such compounds may inhibit FDTs with little effect on classical thymidylate synthases and thus may serve as leads to selective antibiotics that would not interfere with human thymine biosynthesis.

METHODS SUMMARY

Purification and activity of FDTs enzymes. The FDTs from *T. maritima* (TM0449, GenBank accession number NP228259), and its mutants S88A and S88C, were expressed and purified as previously described⁶. The activities of these enzymes were determined using a [2-¹⁴C]dUMP assay which is a modification of the procedure developed and described in ref. 11. Mutant reactivity was also determined by oxidation of chemically reduced enzyme by CH₂H₄folate and dUMP under an atmosphere of purified Ar.

Halogenated substrate derivatives. The 5Br-dUMP assay was adopted from ref. 16. A thymidylate synthase inhibitor, 5F-dUMP, was assessed as a covalent inhibitor of FDTs by incubating it with the enzyme in the presence of CH₂H₄folate and sodium dithionite, followed by removal of small molecules by ultra filtration. Activities of the incubated enzymes were determined before MALDI-TOF analyses.

Mass spectroscopy and NMR analyses. All MALDI-TOF and ESI-MS analyses were conducted at the High Resolution Mass Spectrometry Facility (HRMSF) of the University of Iowa. Enzymes were analysed after trypsin digestion. NMR analyses were performed at the University of Iowa NMR Central Research Facility and Medical NMR Facility using Bruker model Av-300 (for ¹H NMR measurements) and Av-800 (for ²H NMR measurements) spectrometers.

Following the flow of deuterium during the FDTs reaction in D_2O . Experiments were performed at 65 and $37^\circ C$ using dUMP or 6D-dUMP, NADPH, CH₂H₄folate and enzyme in D_2O or H_2O under anaerobic conditions. The product dTMP was then purified using semi-preparative reverse-phase HPLC.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 December 2008; accepted 26 February 2009.

- Carreras, C. W. & Santi, D. V. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**, 721–762 (1995).
- Finer-Moore, J. S., Santi, D. V. & Stroud, R. M. Lessons and conclusions from dissecting the mechanism of a bisubstrate enzyme: thymidylate synthase mutagenesis, function and structure. *Biochemistry* **42**, 248–256 (2003).
- Mylykallio, H. *et al.* An alternative flavin-dependent mechanism of thymidylate synthesis. *Science* **297**, 105–107 (2002).
- Leduc, D. *et al.* Functional evidence for active site location of tetrameric thymidylate synthase X at the interphase of three monomers. *Proc. Natl Acad. Sci. USA* **101**, 7252–7257 (2004).
- Chernyshev, A., Fleischmann, T. & Kohen, A. Thymidyl biosynthesis enzymes as antibiotic targets. *Appl. Microbiol. Biotechnol.* **74**, 282–289 (2007).
- Lesley, S. A. *et al.* Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl Acad. Sci. USA* **99**, 11664–11669 (2002).
- Mathews, I. I. *et al.* Functional analysis of substrate and cofactor complex structures of a thymidylate synthase-complementing protein. *Structure* **11**, 677–690 (2003).
- Leduc, D. *et al.* Two distinct pathways for thymidylate (dTMP) synthesis in (hyper)thermophilic Bacteria and Archaea. *Biochem. Soc. Trans.* **32**, 231–235 (2004).
- Escartin, F., Skouloubris, S., Liebl, U. & Mylykallio, H. Flavin-dependent thymidylate synthase X limits chromosomal DNA replication. *Proc. Natl Acad. Sci. USA* **105**, 9948–9952 (2008).
- Mylykallio, H., Leduc, D., Filee, J. & Liebl, U. Life without dihydrofolate reductase FolA. *Trends Microbiol.* **11**, 220–223 (2003).
- Agrawal, N., Lesley, S. A., Kuhn, P. & Kohen, A. Mechanistic studies of a flavin-dependent thymidylate synthase. *Biochemistry* **43**, 10295–10301 (2004).

12. Graziani, S. *et al.* Catalytic mechanism and structure of viral flavin-dependent thymidylate synthase ThyX. *J. Biol. Chem.* **281**, 24048–24057 (2006).
13. Sampathkumar, P. *et al.* Structure of the *Mycobacterium tuberculosis* flavin dependent thymidylate synthase (MtbThyX) at 2.0 Å resolution. *J. Mol. Biol.* **352**, 1091–1104 (2005).
14. Hong, B., Maley, F. & Kohen, A. The role of Y94 in proton and hydride transfers catalyzed by thymidylate synthase. *Biochemistry* **46**, 14188–14197 (2007).
15. Hyatt, D. C., Maley, F. & Montfort, W. R. Use of strain in a stereospecific catalytic mechanism: crystal structures of *Escherichia coli* thymidylate synthase bound to FdUMP and methylenetetrahydrofolate. *Biochemistry* **36**, 4585–4594 (1997).
16. Wataya, Y. & Santi, D. V. Thymidylate synthase catalyzed dehalogenation of 5-bromo- and 5-iodo-deoxyuridylate. *Biochem. Biophys. Res. Commun.* **67**, 818–823 (1975).
17. Hong, B. & Kohen, A. Microscale synthesis of isotopically labeled 6R-N⁵, N¹⁰ methylene-5, 6, 7, 8-tetrahydrofolate. *J. Labelled Comp. Radiopharm.* **48**, 759–769 (2005).
18. Gattis, S. G. & Palfrey, B. A. Direct observation of the participation of flavin in product formation by thyX-encoded thymidylate synthase. *J. Am. Chem. Soc.* **127**, 832–833 (2005).
19. Klötzer, W. Two isomers of thymine. *Monatsh. Chem.* **104**, 415–420 (1973).
20. Carey, F. A. & Sundberg, R. J. *Advanced Organic Chemistry, Part A* (Kluwer Academic/Plenum, 2000).
21. Brown, B. J., Deng, Z., Karplus, P. A. & Massey, V. On the active site of Old Yellow Enzyme. Role of histidine 191 and asparagine 194. *J. Biol. Chem.* **273**, 32753–32762 (1998).
22. Fagan, R. L., Nelson, M. N., Pagano, P. M. & Palfrey, B. A. Mechanism of flavin reduction in class 2 dihydroorotate dehydrogenases. *Biochemistry* **45**, 14926–14932 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH R01 GM065368 and NSF CHE 0715448 to A.K., the Iowa Center for Biocatalysis and Bioprocessing to E.M.K., NIH R01 GM61087 to B.A.P., NIH training grant GM08270 to J.A.C., and JCSG grant U54GM074898 to S.A.L. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of DOE, OBER. The SSRL Structural Molecular Biology Program is supported by DOE, OBER and by NIH, NCRR, Biomedical Technology Program and NIGMS.

Author Information Atomic coordinates and structure factor files have been deposited with the Protein Data Bank under the accession codes 3g4a and 3g4c. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.K. (amnon-kohen@uiowa.edu).

METHODS

Materials. All materials were reagent grade and used without further purification unless specified. 2'-deoxyuridine-5'-monophosphate (dUMP), reduced nicotinamide adenosine dinucleotide phosphate (NADPH), 5-bromo-2'-deoxyuridine (5Br-dU), 5-fluoro-2'-deoxyuridine-5'-monophosphate (5F-dUMP), trypsin protease, ammonium bicarbonate, tris(hydroxymethyl)aminomethane, D₂O, phosphocreatine and creatine kinase were purchased from Sigma. Radiolabelled [2-¹⁴C]dUMP was obtained from Moravsek Biochemicals. N⁵,N¹⁰-methylene-5,6,7,8-tetrahydrofolate (CH₂H₄folate) was a gift from Eprova Inc. The FDTs from *T. maritima* (TM0449, GenBank accession number NP228259) and its mutants S88A and S88C were expressed and purified as previously described⁶. The thymidine kinase plasmid was obtained from R. Stroud's laboratory at UCSF and expressed and purified as described in ref. 23.

Synthesis of 5-bromo-2'-deoxyuridine-5'-monophosphate (5Br-dUMP). 5Br-dUMP was synthesized by phosphorylation of 5Br-dU at 37 °C in a 100 mM Tris, 10 mM MgCl₂ buffer at pH 7.5. The reaction mixture contained 1.5 mM 5Br-dU, 5 mM ATP, 50 mg ml⁻¹ phosphocreatine, 2 mg ml⁻¹ creatine kinase and ~1 μM of thymidine kinase. The 5Br-dUMP product was purified by HPLC-UV/Vis (after 280 nm absorbance) and analysed by ESI-MS.

Synthesis of dUMP with deuterium at C6. This procedure was adapted from ref. 24. dUMP (225 mg) was dissolved twice in 5 ml of D₂O (>99.96 D atom) under Ar gas, and evaporated under vacuum (<50 mtorr) to dryness to reduce proton contamination. The dUMP was then dissolved in 5 ml of D₂O and stirred in the presence of Pt(IV) oxide under 1 atm D₂ gas (>99.96 D atom) for 3 h. Vacuum filtration removed the catalyst and the remaining solution was lyophilized to dryness. NMR analysis confirmed >99.5% D atom substitution of both the 5 and 6 positions of the uracil ring. A method to substitute the 5D into 5H without affecting 6D has been developed¹⁴; however, such substitution was not used in the current preparation because the thymidylate synthase reaction is a substitution reaction where a methyl group replaces the 5H of dUMP to form dTMP. Because the 5H position is always replaced during the synthesis of dTMP^{1,2,4} its isotopic labelling can be disregarded.

Analytical methods. All analytical separations were performed on an Agilent series HPLC model 1100, equipped with online degasser, UV/Vis diode array detector and 500TR series Packard flow scintillation analyser (FSA). A Supelco reverse phase column (Discovery series 250 mm × 4.6 mm) was used starting with 100 mM KH₂PO₄ (pH 6.0) followed by a methanol gradient as described elsewhere¹¹. The enzyme active site concentration was determined by the 454 nm absorbance of bound FAD ($\epsilon = 11,300 \text{ cm}^{-1} \text{ M}^{-1}$).

Purification methods. Separation by HPLC was performed using a semi-preparative reverse phase Supelco Column (Discovery series 250 mm × 10 mm). Mobile phase used for separation was a gradient of 100 mM KH₂PO₄ and methanol. Eluent was collected according to the ultraviolet spectral absorbance of the purified species and then lyophilized to dryness.

Protein crystallization. The protein-FAD-dUMP complex was prepared by treating 15 mg ml⁻¹ of the enzyme with around 10 molar excess of dUMP. The well solution for crystallization contained 35–45% PEG200 and 0.1 M Tris-HCl (pH 8.0) buffer.

Activity assays. The activity assay ([2-¹⁴C]dUMP assay) was a modification of the procedure developed and described in ref. 11. All experiments were performed in 200 mM Tris buffer (exchanged with Ar) pH 8.0 at 65 °C with standard reaction conditions of: 100 μM dUMP (including 0.5 mega-decompositions per min [2-¹⁴C]dUMP), 200 μM CH₂H₄folate, 5 mM CH₂O (to stabilize CH₂H₄folate) and 5 mM sodium dithionite. Reactions were initiated by addition of 0.1–2 μM (final concentration) of enzyme, quenched with HCl (to a final pH of 1) and

stored at –80 °C until analysis. HPLC-FSA analysis was used to determine the conversion of [2-¹⁴C]dUMP to [2-¹⁴C]dTMP.

5Br-dUMP assay. Reactions with either *E. coli* thymidylate synthase or *Tm*FDTs were performed in a 200 mM Tris buffer at pH 8.0, containing 100 μM 5Br-dUMP, 5 mM sodium dithionite, and 5 mM β-mercaptoethanol. Enzyme was added to the reaction mixture which was then incubated at 37 °C for 3 h. Product conversion was determined by HPLC-UV/Vis analysis of the reaction mixtures by following both 256 and 280 nm absorptions for dUMP and 5Br-dUMP, respectively.

Oxidative half-reaction of S88A. A solution of oxidized S88A (20 μM) and dUMP (1 mM) was made anaerobic in a sealed cuvette by successive cycles of evacuation and equilibration with an atmosphere of purified Ar. The oxidized enzyme was titrated spectrophotometrically to complete reduction with a solution of dithionite. CH₂H₄folate was added anaerobically from a side arm to initiate the reaction (25 °C). The absorbance spectrum of oxidized enzyme returned before the first scan, indicating a rapid reaction of the mutant enzyme.

Protein digestion. All enzyme digestion reactions were performed in 100 mM ammonium bicarbonate buffer adjusted to pH 8.0 at 37 °C. Enzyme solutions were diluted to 1 μM protein concentration followed by addition of trypsin to a final concentration of 10 ng μl⁻¹. All digestions were allowed to incubate for 3 h at 37 °C and stored at –20 °C before MALDI-TOF mass spectrometry analysis.

Assessment of 5F-dUMP as a covalent inhibitor of FDTs. These experiments were performed at 37 °C in 200 mM Tris buffer at pH 8.00. Both wild-type FDTs and S88A (11 μM active site concentration) were incubated for 30 min in the presence of 50 μM 5F-dUMP, 200 μM CH₂H₄folate and 5 mM sodium dithionite. After incubation, activities of the enzymes were determined using the standard activity assay conditions (except with a residual 5 μM 5F-dUMP). Samples of native and trypsin-digested enzymes were prepared for MALDI-TOF MS analysis. The remaining enzyme solutions were exchanged with 40 ml of Tris buffer at 4 °C, and concentrated by centrifugal filtration (using a Millipore 10,000 MWCO filtration device) to 11 μM active site concentration. Once concentrated, the activity of both FDTs and S88A was determined using the standard activity assay conditions. As described in the report, no covalent adduct of 5F-dUMP to enzyme was identified and both enzymes recovered 100% activity after the removal of the 5F-dUMP from the reaction mixture.

Following the flow of deuterium during the FDTs reaction. For studies in D₂O, all substrates in 100 mM Tris buffer were exchanged twice by dissolving in D₂O (>99.96% D atom) and lyophilizing to dryness before use. Experiments were performed in 100 mM Tris buffer (99.96% D₂O or H₂O) pH 8.0 at 65 and 37 °C using 4 mM dUMP or 6D-dUMP, 8 mM NADPH and 8 mM CH₂H₄folate, under Ar. To maintain anaerobic conditions 10 mM glucose and 100 units of glucose oxidase were added to the reaction mixture. Reactions were initiated by adding enzyme (previously lyophilized and resuspended in D₂O or H₂O) to a final concentration of 1–10 μM. The reaction mixtures were incubated (at 65 or 37 °C) for 20 h under Ar and stored at –20 °C. The product dTMP was then purified using semi-preparative HPLC, lyophilized, triturated into methanol, and dried under vacuum. The dTMP was dissolved in D₂O or H₂O for ¹H and ²H NMR analysis, respectively.

23. Waldman, A. S., Haeusslein, E. & Milman, G. Purification and characterization of herpes simplex virus (type 1) thymidine kinase produced in *Escherichia coli* by a high efficiency expression plasmid utilizing a lambda PL promoter and cl857 temperature-sensitive repressor. *J. Biol. Chem.* **258**, 11571–11575 (1983).
24. Burdzy, A., Noyes, K. T., Valinluck, V. & Sowers, L. C. Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA. *Nucleic Acids Res.* **30**, 4068–4074 (2002).

How to get ahead in imaging

Advances in magnetic resonance imaging are helping scientists learn more about the structure and function of the brain. Nathan Blow looks at how far the technology has developed and where it could go.

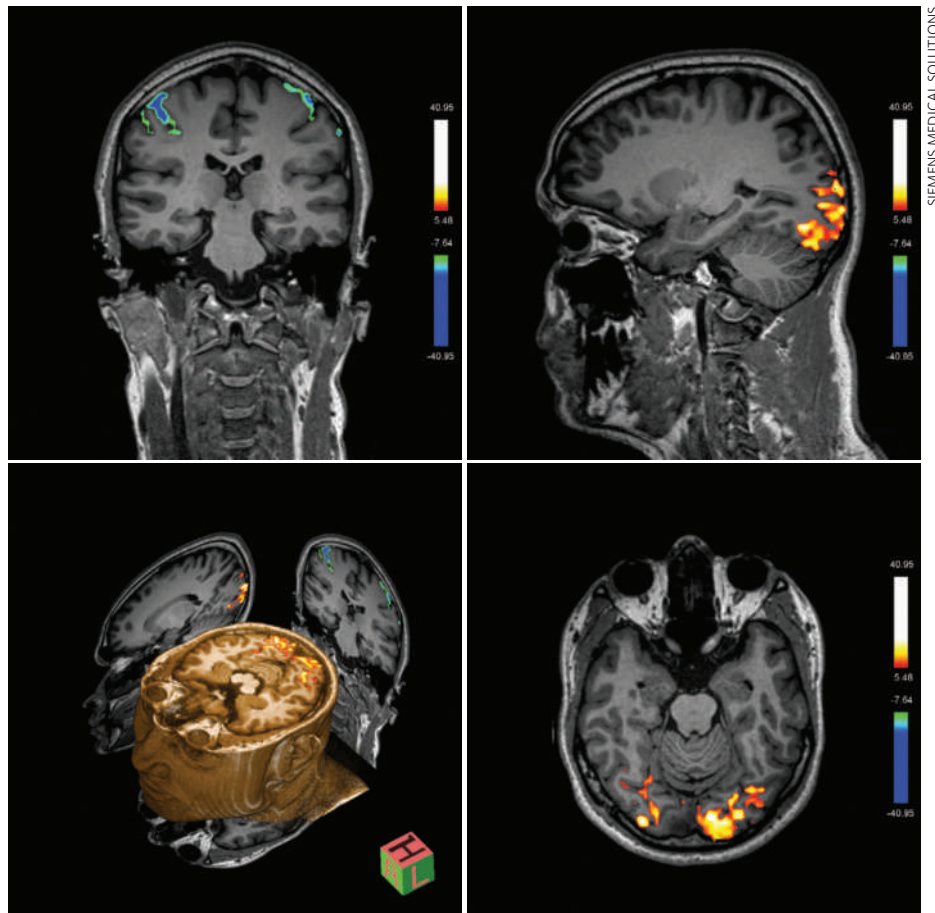
A wave of enthusiasm seems to overcome Joy Hirsch, a neuroscientist at Columbia University in New York, when she talks about recent developments in functional neuroscience. “I just love the new area of neuroeconomics,” she says. Although not her primary field of study — her group focuses on how the brain uses cognitive control in decision-making processes — she says that understanding the neurophysiology behind financial decisions could make it possible to one day predict a person’s financial habits by imaging the neural activity in their brain. And it is advances in cognitive research such as this that could eventually change the way we think about the brain–mind interface, says Hirsch.

The emergence of these areas in functional neuroscience can be traced in part to recent developments in brain imaging technology. “Magnetic resonance imaging used to be only in the hands of radiologists, but today more and more cognitive neuroscientists are using it,” says Alan Koretsky, scientific director of the National Institute of Neurological Disorders and Stroke at the National Institutes of Health (NIH) in Bethesda, Maryland. And by taking advantage of magnetic resonance imaging (MRI), sometimes in combination with other methods such as electroencephalography (EEG) or positron emission tomography (PET), neuroscientists are exploring the structure of the human brain in greater detail than ever before and identifying specific regions that are active when a person is making a decision or feeling an emotion such as fear or happiness.

Size is important

With its high sensitivity and non-invasive nature, MRI is at the core of neuroimaging today. To create detailed anatomical and functional images, MRI systems take advantage of the ability of very large cylindrical magnets, ranging in strength from 1.5 to 15 tesla, to align the protons found in water throughout the body. Smaller, localized radio-frequency electromagnetic fields are then generated to push those protons out of alignment. The displaced protons generate signals that are detected by the MRI instrument and translated into an image.

Whether built for imaging humans or small animals, MRI instruments are defined by the field strengths of their magnets — and here bigger is better. The signal from magnetic resonance is inherently weak and it can be



Magnetic resonance images, such as these taken with a 7-tesla scanner, are getting more and more detailed.

difficult to detect, which tends to limit resolution. So developers are constantly trying to increase the field strengths of their magnets and boost the signal.

“Small-animal imaging is based on making systems with magnets of smaller diameter but much higher field strength than you would get in clinical practice,” says Roy Gordon, vice-president of imaging at Bruker Biospin, a scientific-instrumentation company in Billerica, Massachusetts. Most animal-imaging magnets range in field strength from 4.7 T to 15 T, with 7 T being the “work-horse field strength in animal imaging,” according to Gordon. Human MRI instruments, on the other hand, use 0.5–3 T magnets

for clinical applications and up to 9.4 T for research applications.

Part of the challenge in moving human scanners up in field strength to match their small-animal counterparts comes from

our body sizes. “Maintaining field strength homogeneously over a large volume becomes more and more challenging as the field strength increases,” says Vibhas Deshpande, a research and development scientist at Siemens Medical Solutions in Malvern, Pennsylvania. For this reason, he says, the human scanners that are operating at ultra-high field strengths are mostly using smaller diameter magnets, similar to animal scanners, to image small human samples such as tissues rather than performing whole-body scans.

Image artefacts

Maintaining homogenous field strength is not the only issue that magnet builders face in their quest to boost the signal. “Another challenge for all systems is that as you go up in field strength, you must address issues related to magnetic susceptibility,” says Gordon. At higher field strengths, a magnetic field gradient can occur at the interface of materials with very different magnetic susceptibilities, such

“We would love an anatomical picture that looks like a histological slice.”

— Alan Koretsky

as tissue, bone and a void (in the sinuses). This can lead to artefacts in the images that must be accounted for during image analysis, says Gordon.

Currently, 3 T MRI systems are the standard for high-end human neuroimaging. But research-grade 7 T MRI instruments for human studies have come a long way in recent years with several companies, including Siemens Medical Solutions, GE Healthcare in Piscataway, New Jersey, and Philips in Andover, Massachusetts, now supplying second-generation versions of these systems. "The first generation of 7 T systems were monsters that needed 400 tonnes of shielding, but the second-generation systems are actively shielded so now in principle you don't need any iron to shield the magnet," says Gordon. But many researchers say that, even with the advances, 7 T systems still need more engineering work.

The 'best' field strength?

"The 7 T scanner is still a bit of a specialized device and, in my opinion, has yet to reach its full potential," says Larry Wald, director of the NMR core facility at the Athinoula A. Martinos Center for Biomedical Imaging in Charlestown, Massachusetts. Deshpande agrees: "There will be some ramp-up time with



The radio-frequency coils used to transmit and detect MRI signals are being made more sensitive.

the 7 T in terms of applications and research. We need to find out how far we can push the instrument."

Whether 7 T will eventually become a clinically robust field strength for human MRI is not clear. "Ever since the beginning of MRI there has always been discussion about what field one should work at," says Koretsky. First there was the decision between using 0.5 T and 1.5 T, and then between 3 T and 4 T magnets. Now, he says, a 7 T debate might start. "Some of the images from the brain at 7 T are truly amazing," says Deshpande. "We are seeing things that we have never seen at either 1.5 T or 3 T."

Even as new 7 T human systems and 11.7 T and higher animal systems from companies

such as Bruker Biospin and Varian in Palo Alto, California, expand in use within the neuroscience community, bigger magnets are being designed for cutting-edge animal and human imaging. In autumn 2008, the Martinos Center installed a 15 T magnet designed by Varian and Magnex Scientific in Walnut Creek, California, for imaging mice and rats, which Wald says is now at field strength and should be generating its first images in the coming weeks. The machine, used in conjunction with knockout mice with genes that have been 'turned off', should allow scientists to understand neural disease progression more effectively and even test potential drug therapies, says Wald.

Field strengths for human MRI magnets may reach double digits in the coming years. Both the NIH and NeuroSpin, a centre for ultra-high-field MRI in Saclay, France, have announced projects to construct 11.7 T MRI magnets for imaging human subjects. Although Koretsky says that it will be several years before these new magnets are installed and operational, he thinks the images will be worth the wait. "In the end you would love an anatomical picture that looks like a histological slice," he says, adding that these new magnets will get them closer than ever to making this a reality.

L. WALD

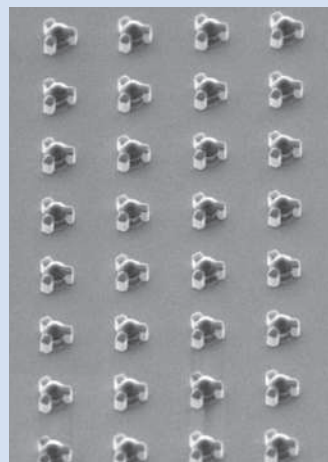
CHANGING THE COLOUR OF MRI

When Alan Koretsky, scientific director of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, and Gary Zabow began to think about developing new contrast agents for magnetic resonance imaging (MRI), they took their design cues from the colourful world of molecular imaging. "It was looking at what existed and then figuring out some way to copy the idea of quantum dots," says Zabow, a physicist at the National Institute of Standards and Technology in Boulder, Colorado.

In molecular imaging, a quantum dot can generate a range of possible emission spectra simply by varying the size of the dot's inner core shell. This is a stark contrast to the traditional agents used in MRI, such as gadolinium or iron oxide, which are magnetic materials that alter the signal from the protons in the surrounding water, appearing as either darker or brighter spots on images. "It is sometimes difficult to tell the different agents apart from one another or from artefacts that make something brighter or darker," says Zabow. (At present, colour in MRI scans — such as

those in this article — is assigned to shades of grey and added during processing.)

"One of the areas of MRI that has exploded over the past 5 or 6 years is the ability to track cells as they move around," says Koretsky. Although MRI cannot achieve single-cell resolution, a single cell can have sufficient magnetic-resonance contrast to be detected. But for researchers interested in tracking two or three cells at once, this level of differentiation is not enough.

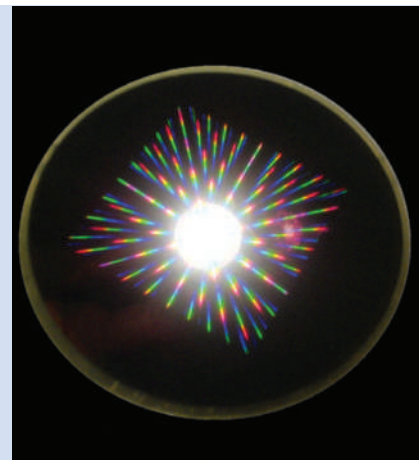


Electron-microscopy images of microfabricated contrast agents.

So Zabow and Koretsky microfabricated specific magnetic shapes that would create different magnetic fields and so shift the nuclear magnetic resonance frequency. "The existing magnetic particles do not shift the frequency — they just broaden it out," says Zabow. But the very precise shape of these new agents generates a corresponding precise frequency shift, similar to quantum dots, giving Zabow and Koretsky the possibility of creating different colours through different shapes and their specific frequency shifts.

The initial work consists of two discs with a gap between them in which the magnetic field can be generated. By varying the thickness or diameter of the discs, or the gap, different fields can be obtained so that when water passes between the discs the magnetic resonance of the water molecules flowing through the gap shifts (G. Zabow *et al. Nature* 453, 1058–1063; 2008).

Although the possible range of new colours is still to be determined, that is not the primary focus at the moment, says Zabow. They are working to improve the



Microfabricated shapes provide potential for a wide range of possible colours when used in MRI.

fabrication process and make the magnetic particles smaller and more robust. "We are working first on that. The idea of having as many colours as possible falls out from there because in improving the fabrication we are getting the geometry more precise," he says.

Koretsky sees these new contrasting agents as adding a unique ability to MRI that no other radiological imaging technique possesses.

G. ZABOW

N.B.

High-strength magnets come with big technical challenges and hefty price tags. “These are demanding instruments. We can manufacture 16 T magnets for small-animal imaging but they are exceptionally expensive,” says Gordon. With a complete imaging system based on a 16 T magnet for small-animal imaging costing tens of millions of dollars and magnets for imaging humans often costing similar amounts, Gordon and others think that advances in radio-frequency-coil design could provide a more economical route to higher sensitivity.

“It is easy to forget, but if you dust off the old birdcage detectors that we used to use for the human head and compare those to what is shipping on a modern scanner these days — the sensitivity difference is pretty astonishing,” says Wald. His group at the Martinos Center has been developing the coils for many years, but it might be the group’s contributions to the development of parallel imaging that have had the biggest impact for MRI.

Parallel evolution

Radio-frequency coils are used as both transmitters to oscillate protons and as detectors to receive the signal. Ten years ago, single-channel coils were used for this purpose. But today, most instruments have multi-channel coils, allowing the parallel acquisition of data during a scan. The benefits of parallel imaging can be seen in the 32-channel coil for head scans that Wald’s group developed for brain imaging on Siemens’ research-grade 7 T human MRI system. “We have a protocol that used to run as two 8.5-minute scans, but now we can do that in a single 3.5-minute scan,” he says. Faster scans mean less time in the instrument for subjects, which presents new opportunities for the researchers at the larger imaging centres who routinely scan

thousands of people a year. “There are now a lot of scientists interested in standardizing a fast morphometric protocol that would enable large-scale genotype/phenotype studies of neurological diseases,” says Wald.

Wald’s group has also worked at getting a better signal from its scans at lower field strengths with the development of a 96-channel head coil for use on Siemens’ clinical 3 T MAGNETOM system. Although using more channels should mean better signal detection, this is not always the case. “In principle, if it is a 96-channel coil you can get a 96 factor acceleration in scan speed, but we do not come anywhere close to that,” says Wald, a fact that is leading coil groups around the world to work on further

enhancing the sensitivity of multi-channel coils.

The developments are leading to advances in commercially available coils. With its total imaging matrix (TIM) technology for scanning the whole body, Siemens offers multiple coils that can connect at the same time, allowing the system to decide on which of the coils to use depending on the section of body being scanned. In this way larger sections can initially be scanned at low resolution, followed by more focused higher-resolution scans of sections of interest, all without having to change the coil setups or move the patient out of the scanner to reposition the coil array. Philips now offers up to 33-element coils for neuroimaging on its 3 T Achieva TX system and GE Healthcare offers a series of coils designed for its Signa Infinity 1.5 T system.

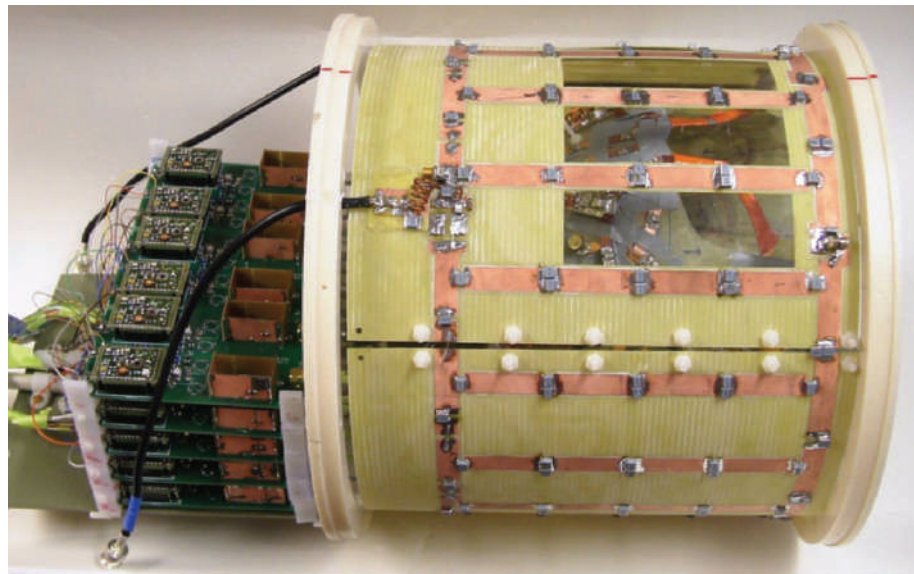
Divide and conquer

As magnet field strengths increase, coil developers will have to keep improving sensitivity. “The detection side of it will continue to be hard, but we know what to do and how to do it now. The transmit side on the other hand is potentially most challenging,” says Koretsky.

For radio-frequency transmitting, the technical challenges increase with field strength because the human body starts to distort the radio frequency at high field strengths affecting the radio-frequency transmission. Some progress has been made on this front already, with researchers working on the idea of parallel transmission using multiple radio-frequency transmitters that are simultaneously excited. By using parallel transmission it is possible to eliminate the homogeneity problems caused by the shorter wavelengths generated



Siemens’ MAGNETOM Verio 3 T scanner has a head and neck coil detection system.



Multi-channel coils can reduce the time that patients need to stay in an MRI scanner.

at higher frequencies. “Parallel transmit breaks it down into chunks, which can be put back together to get a uniform transmit in the end,” says Wald. “A sort of ‘divide and conquer’ approach helps to get around this.”

But another recent development might further change how transmission is done. In February 2009, Klaas Pruessmann’s group at the Institute for Biomedical Engineering in Zurich, Switzerland, described a new transmission and detection technique called travelling-wave nuclear magnetic resonance (D. O. Brunner *et al. Nature* 457, 994–998; 2009). In this approach, the radio-frequency wave travels down the bore of the magnet, which acts like a waveguide. Although this means that the bore has to be big enough compared with the wavelength for the wave to travel, many researchers are encouraged by the potential of this approach to solve some of the radio-frequency transmission issues.

“The most exciting thing with the travelling wave approach is that you not only get the apparatus for transmit out of the bore of the magnet, but it also offers the potential to have a more homogenous excitation,” says Wald. Although the approach will only work for 7 T-and-higher strength systems, Wald says that this is where this type of technology is most needed, given the transmission problem and the decreasing bore sizes that come with larger magnets.

Adding function to the form

Signal in MRI is much like currency, says Deshpande. By boosting the signal with stronger magnets and multi-channel coils, researchers can “spread their currency around to get shorter scan times and higher

resolution when doing functional imaging”.

Functional MRI (fMRI) is a technique that looks for changes in the levels of deoxyhaemoglobin — used as an internal contrast agent (see ‘Changing the colour of MRI’, page 926) — as an indicator of blood flow in the brain due to neural activity. By observing neural activity in subjects performing different tasks, fMRI is helping cognitive neuroscientists gain a better handle on how the brain works. “It is that intimate relationship between structure and function through the melding of behaviour science and experimental paradigms along with the ability to image the brain while doing these experiments that grows the field,” says Hirsch.

Although the field continues to grow through technology development and improving experimental design, mainstream applications of fMRI remain limited. Neurosurgeons are using the technique to map patients’ brains before surgery, but at the

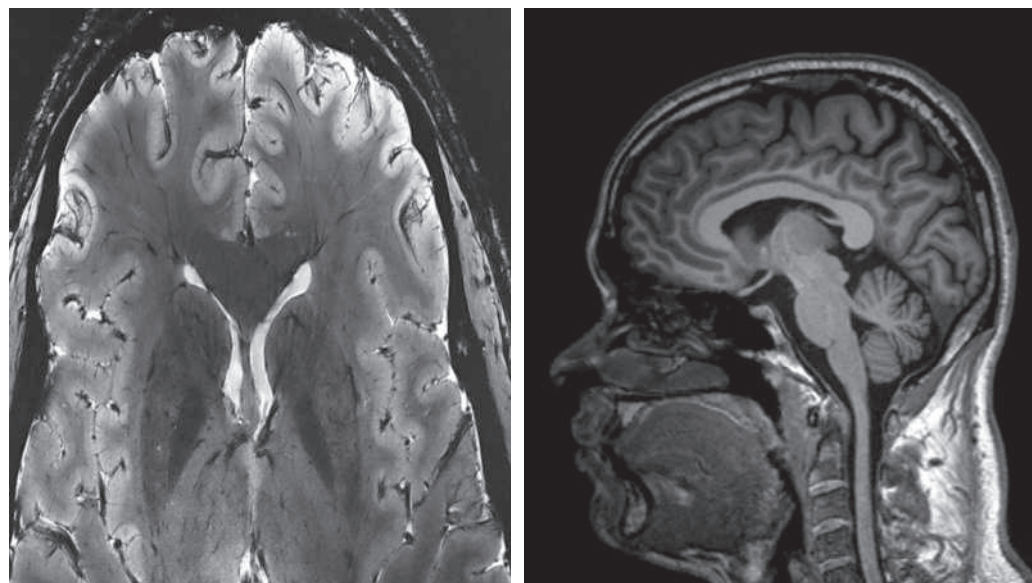
moment this is the only application of fMRI for which physicians can be reimbursed through insurance companies in the United States.

But this is not stopping researchers from moving forwards with new fMRI studies. Omneuron, a life-sciences company in Menlo Park, California, is currently recruiting volunteers for a clinical trial on the neural basis of chronic pain. Using fMRI, researchers at Omneuron are mapping the brain activities of participants performing specific mental tasks. Other companies, such as the Applied fMRI Institute in San Diego, California, with its 3 T Siemens Tim MAGNETOM Trio scanner, provide both fMRI and anatomical MRI services, and a host of companies, including Compumedics Neuroscan in Charlotte, North Carolina, Visage Imaging in Andover, Massachusetts, and Prism Clinical Imaging in Wauwatosa, Wisconsin, are providing software to integrate the functional information of fMRI with other modalities including EEG and PET.

The biggest changes in fMRI might come as the technology becomes more personalized. “One of the signs of brain imaging advancement is how well can we apply the benefits of neuroimaging to personalized patient care,” Hirsch says. But even more sensitivity and accuracy will probably be required for this, because fMRI studies often rely on multiple subjects.

By improving understanding of why we make certain decisions and how our brains differ in shape, MRI is changing the way we view the brain. And many researchers think more insights are on the way as the technology continues to advance. “MRI is going to move from a tool for identification to one that helps us understand mechanisms,” says Koretsky. “From being able to say there is a tumour there to understanding how that tumour formed and grew.”

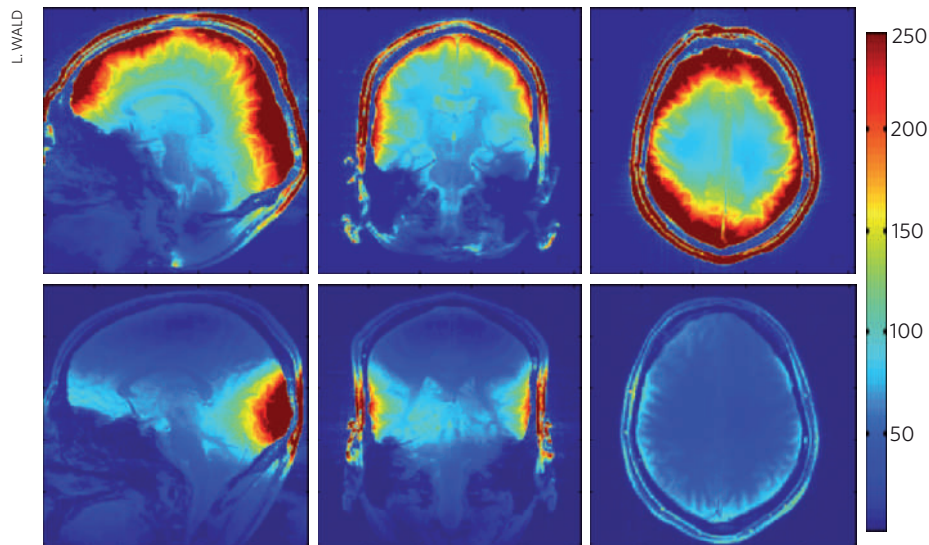
Nathan Blow is the technology editor for *Nature* and *Nature Methods*.



Images of the human brain taken with 7-tesla MRI scanners.

SIEMENS MEDICAL SOLUTIONS, L. WALD

OPTIMUM SIGNAL-TO-NOISE RATIO



The MRI signal with a 32-channel coil array (top row) is clearer than with an 8-channel array (bottom row).

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
MRI, MEG, PET and EEG instrumentation			
Applied fMRI Institute	fMRI and anatomical MRI services	San Diego, California	www.appliedfmri.org
Bioscan	PET, MRI and SPECT for small animals	Washington DC	www.bioscan.com
Bruker BioSpin	Small-animal MRI imaging systems and coils	Billerica, Massachusetts	www.bruker-biospin.com
Elekta	MEG instrumentation for functional mapping of human brain; gamma-knife technology development	Stockholm, Sweden	www.elekta.com
GE Healthcare	Human clinical and research-grade MRI systems and coils	Chalfont St Giles, UK	www.gehealthcare.com
Hitachi Medical Systems	Oasis 1.2 T and Echelon 1.5 T clinical imaging systems and coils	Tokyo, Japan	www.hitachimed.com
InnerVision MRI	Specialized MRI systems for whole body, orthopaedic and neonatal imaging	London, UK	www.innervision-mri.co.uk
Philips Healthcare	Health-care instrumentation including MRI	Andover, Massachusetts	www.medical.philips.com
Siemens Healthcare	MRI and RF coils	Munich, Germany	www.medical.siemens.com
Toshiba America MRI	Vantage 1.5 T MRI system; computerized tomography and ultrasound	Tustin, California	medical.toshiba.com
Varian Medical Systems	Small-animal MRI systems and technology	Palo Alto, California	www.varian.com
MRI accessories			
Applied Science Laboratories	Eye-Trac6 system for long-range eye tracking during fMRI experiments	Bedford, Massachusetts	www.a-s-l.com
Carolina Medical Parts	Sales and repair of MRI coils and RF amplifiers	Winston Salem, North Carolina	www.cmparts.com
m2m imaging	Cryogenic, phased array and conventional coils for clinical and preclinical applications; imaging accessories for PET, SPECT and CT	Cleveland, Ohio	www.m2mimaging.com
Microsemi	Integrated circuits and devices for the medical field; RF products	Irvine, California	www.microsemi.com
Midwest RF	Array coils for head, brain and spine; custom coil development	Hartland, Wisconsin	www.midwestrf.com
MRlequip.com	Accessories for clinical MRI instruments	Nisswa, Minnesota	www.mriequip.com
NordicNeuroLab	fMRI solutions and software; fMRI hardware and accessories; MR simulator	Bergen, Norway	www.nordicneurolab.com
Resonance Technology	Audio and visual devices for use with fMRI and clinical MRI applications	Northridge, California	www.mrvideo.com
ScanMed	Coils for MRI for various locations of the body; repair, manufacturing and refurbishment of MRI coils	Omaha, Nebraska	www.scanmed.com
T2 Biosystems	Miniaturized MR technology using nanotechnology	Cambridge, Massachusetts	www.t2biosystems.com
Zetta Medical Technologies	Service and repair of MRI instruments, coils and magnets; computer-tomography instrument repair services	Lake Zurich, Illinois	www.zettamed.com
Imaging Software			
Advanced Neuro Technology	Software for the integration of fMRI data with EEG, MEG and ERP recordings	Enschede, the Netherlands	www.ant-neuro.com
Brain Innovation	Advanced analysis and visualization tools for structural and fMRI data	Maastricht, the Netherlands	www.brainvoyager.com
Compumedics	Software platforms for the integration of fMRI data with EEG, PET and anatomical imaging	Abbotsford, Australia	www.compumedics.com
Medical Numerics	fMRI analysis software package; tools for functional and structural brain-image analysis	Germantown, Maryland	medx.sensor.com
MRVision	Tools for visualization and analysis of MRI, CT, fMRI and diffusion mapping data	Winchester, Massachusetts	www.mrvision.com
Prism Clinical Imaging	Software packages for the visualization of MRI and fMRI data	Wauwatosa, Wisconsin	www.prismclinical.com
TomoVision	sliceOmatic software package for the computation of anatomical volumes from MRI and CT images	Montreal, Canada	www.tomovision.com
VayTek	Deconvolution software, 3D volume visualization	Fairfield, Iowa	www.vaytek.com
Visage Imaging	MRI visualization software; 3D reconstructions	Andover, Massachusetts	www.visageimaging.com
Visiopharm	Software packages for analysis and management of imaging data	Hørsholm, Denmark	www.visiopharm.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
General			
ALEXIS Biochemicals	Reagents for molecular and cell-biology research	Lausen, Switzerland	www.alexis-corp.com
Cambrex	Molecular and cell-biology reagents and tools	East Rutherford, New Jersey	www.cambrex.com
Carl Zeiss	Microscopes, digital imaging of cells, image-analysis software	Jena, Germany	www.zeiss.com ●
Chroma Technology	Optical filters	Rockingham, Vermont	www.chroma.com
Clontech	Reagents for PCR, nucleic-acid extraction and purification; RNAi expression systems	Mountain View, California	www.clontech.com
EMD (Novagen)	Reagents and kits for molecular biology	Madison, Wisconsin	www.emdbiosciences.com
Epibio	Enzymes for PCR and RT-PCR; DNA and RNA purification and isolation kits	Madison, Wisconsin	www.epibio.com
Eppendorf	Consumables for molecular biology; instrumentation	Hamburg, Germany	www.eppendorf.com
Evrogen	Enrichment kits for transcriptomic applications; fluorescent probe development	Moscow, Russia	www.evrogen.com
Gilson	Pipettes, automated liquid handling, liquid-chromatography systems and software	Middleton, Wisconsin	www.gilson.com
Hamamatsu	Imaging systems	Hamamatsu City, Japan	www.hamamatsu.com
Hamilton Company	Robotics for life sciences	Reno, Nevada	www.hamiltoncomp.com
HORIBA Scientific	Spectroscopy systems and accessories including Raman, atomic-emission and UV spectroscopy	Kyoto, Japan	www.horiba.com/scientific
Invitrogen	Kits and reagents for molecular-biology, genomics and cell-biology research	Carlsbad, California	www.invitrogen.com
Lonza	Molecular-biology reagents and systems	Basel, Switzerland	www.lonza.com ●
Maxim Biotech	PCR reagents and systems; nucleic-acid isolation; custom services including primer design, sample preparation and library construction	Rockville, Maryland	www.maximbio.com
Merck	Chemicals, kits and reagents for molecular- and cell-biology-related research	Darmstadt, Germany	www.merck.de
MP Biomedicals	Reagents and chemicals	Irvine, California	www.mpbio.com
New England Biolabs	Molecular-biology-related reagents, kits and enzymes	Ipswich, Massachusetts	www.neb.com ●
Nikon Instruments	Optical microscopy systems and analysis tools	Tokyo, Japan	www.nikon.com
Promega	Molecular-biology and proteomics equipment and kits	Madison, Wisconsin	www.promega.com
Roche Diagnostics	Reagents and kits for molecular biology; genomics instrumentation and software	Basel, Switzerland	www.roche-applied-science.com
Semrock	Optical filters for microscopy	Rochester, New York	www.semrock.com
Stratagene	Tools for molecular-biology, proteomics and genomics research	La Jolla, California	www.stratagene.com
Takara Bio	Reagents, kits and services for genomics and molecular-biology research	Shiga, Japan	www.takara-bio.com
Thermo Fisher Scientific	Research equipment, chemicals and kits	Waltham, Massachusetts	www.fishersci.com
Tocris Bioscience	Chemicals for life-science research; contract research services	Bristol, UK	www.tocris.com
Wako Chemicals USA	Speciality chemicals	Richmond, Virginia	www.wakousa.com
USB	Chemicals and reagents for molecular biology	Cleveland, Ohio	www.usbweb.com
VisEn Medical	Animal-imaging systems; tomography imaging systems	Bedford, Massachusetts	www.visenmedical.com

● see advertisement

NEWS

Wales woos global graduates

UNIV. WALES

Valuable new scholarships are on offer to lure some of the world's most talented graduate students to Wales. Based at the University of Wales, the students will work with Welsh businesses on innovative products, ranging from nanotechnology devices to aerospace components.

Over the next three years, the £11.4-million (US\$16.8-million) Prince of Wales Innovation Scholarships programme will provide 100 students with an annual stipend of £20,000, a research grant of £5,000 and a tuition waiver. In return, the students are expected to generate new Wales-based products, patents or services.

When the gross domestic product per capita for Wales dipped below 75% of the European Union (EU) average in 1999, it qualified for funds from the European Regional Development Fund, an EU initiative to strengthen economic growth in struggling areas. From a second round of funding in 2006, University of Wales vice-chancellor Marc Clement combined £5 million in development funds with core university funds to promote a knowledge economy focused on innovation.

By embedding the students in industry, Clement wants to provide an opportunity for young researchers to see first-hand how their ideas can have practical results. A statement by Welsh venture capitalist Michael Moritz caught his ear — that the founders of some of the most successful companies are younger



Marc Clement: partnerships.

than 26, non-American, with backgrounds in science, technology or medicine.

"UK scholarships are often limited to UK or EU applicants. These scholarships allow anybody in the world to come to Wales to study," says Steve Conlan, co-director of Swansea University's Centre for NanoHealth. Conlan says that Swansea's School of Medicine is actively working

to take advantage of the scholarship scheme. "We don't have coal mining any more, and little heavy industry, so we are looking for high-tech, high-value industries to sustain our economic future," he says.

Clement is forging partnerships with companies and universities around the world. He says the scholarships will focus on building synergies between academia and companies in priority sectors defined by the Welsh Assembly government, such as life sciences, engineering, information and communication technologies, and environmental technologies. He has met representatives of top research universities to discuss collaborations.

Steven Beckwith, vice-president for research and graduate studies at the University of California, says the university will advertise the programme to its students. "Time will tell if it will work," he says. "But it is an intriguing way to teach graduate students how the new knowledge they create in an academic setting can be applied to industry."

Virginia Gewin

POSTDOC JOURNAL

Presentation dreams

I can't sit still. The speaker explains something, shines a vibrant green dot on a screen scarred with graphs and data points, but my mind can't grasp her words. My left leg is bouncing. Stuttering applause spreads across the room. As in a dream, I walk behind the podium and face the crowd.

My first carefully rehearsed words tumble over the upturned faces. Others follow, efficient, eager. An excited choreographer on opening night, I watch the words dance for a moment, and then disappear off stage right. I am relaxed, enjoying myself. Even

the sleeping man whose head leans backward, mouth open, does not dampen my elation.

I hear the applause and walk offstage, able now to follow the remaining talks. The crowd filters out, leaving clumps of questioners around the speakers. My co-author shakes my hand and says I did a good job. A moment later he is cornered by a grey-haired professor who slides in front of me. Left alone, I stand awkwardly, looking at groups congregated around the other speakers. I shoulder my backpack and leave the room.

Conference presentations

should be an opportunity to make a name for myself. Instead, each one is a parabola of nervousness, elation and disappointment. But maybe, if I keep working, next time I'll nucleate my own crowd. I imagine that this will help reviewers and search committees recognize my name. Hence, my papers will be noticed, interviews will materialize, and I'll be that much closer to my dream of a professorship.

Sam Walcott is a postdoc in theoretical biophysics at Johns Hopkins University in Baltimore, Maryland.



IN BRIEF

Applications slowing

International graduate applications to US institutes for 2009 increased at a lower rate than in previous years, according to a 7 April report from the US Council of Graduate Schools (CGS).

"We are seeing continuation of trend," says Nathan Bell, CGS director of research and policy analysis. The data, based on 400,000 applications collected this year, should reflect initial effects of the global recession, says Bell. Applications were up by 4%, compared with 6% last year. Those from India and South Korea fell by 9% and 7%, respectively, but those from China rose by 16%. Global applications to physical sciences and engineering programmes slowed to a 4% gain after 7% last year.

Applications in life sciences fell by 2% after gaining 3% last year. This could be a one-time occurrence or part of a larger trend, Bell says.

More PhDs for minorities

The Alliances for Graduate Education and the Professoriate (AGEP), a programme funded by the US National Science Foundation (NSF), has helped increase the number of underrepresented minorities receiving PhDs in science, according to an analysis by the American Association for the Advancement of Science (AAAS).

Although the most recent NSF data show no discernible increase in the overall number of minority PhDs, the AAAS analysis of 66 AGEP institutions found a 33.9% increase between 2001 and 2008, with a 50% increase in natural sciences and engineering. Report co-author Yolanda George, deputy director of the AAAS education and human resources programmes, credits institutions for focusing on academic support and paying greater attention to recruiting and tracking minorities.

Responsible research

Conducting research responsibly is the focus of a new guide from the US National Academies. Aimed at early-career scientists and their mentors, *On Being a Scientist: A Guide to Responsible Conduct in Research* addresses such questions as how to allocate credit for a discovery among a team, how to respond to errors in published works, and how to recognize conflicts of interest that could influence study results. Examples include some recent real-world cases of misconduct.

Steady breeze

Despite the economic downturn, wind energy should be fertile ground for jobseeking engineers and scientists, reports **Amanda Leigh Mascarelli**.

In April 2008, Kristian Dixon graduated from the Delft University of Technology in the Netherlands with a master's diploma in aerospace engineering, specializing in wind energy. He had his pick of several positions in the wind industry. After receiving offers from two major turbine makers, a well-known blades manufacturer and a smaller company specializing in urban wind turbines, Dixon chose a position as a development engineer designing the next generation of turbine blades for Siemens Wind Power in Denmark. "I was very lucky to graduate before the financial crisis had really hit," says Dixon. "The number of offers I got was really only limited by the number of resumé's I sent out and the interviews I was willing to do."

In the Netherlands alone, the number of wind-related research openings ranged from 100 to 200 in the two years before the recession. Now many of those positions have temporarily been put on hold — although there's plenty of potential in the long term. After decades of research and development, and buoyed by ambitious renewable-energy targets throughout the world, the wind industry has become cost competitive with other energy sectors. Projected growth should ensure healthy longer-term job prospects, as governments remain interested in big investments, and fledgling scientists and engineers stay interested in training.

From 2005 to 2008, the wind industry grew rapidly, with the global market for turbines increasing at an average of 30–40% per year. The number of global wind installations more than doubled from 2005 to 2008, enabling a cumulative output of 121,000 megawatts. In the same period, the number of jobs worldwide doubled, with the industry now supporting 440,000 positions.

"In an environment where we'd been close to having full employment in the economy, that was a huge challenge," says Christian Kjaer, chief executive of the European Wind Energy Association. With the economic downturn, he says, finding engineers and technical people is easier.

Policies in Europe suggest that the

long-term demand for wind power will persist. In December 2008, the European Union (EU) adopted targets for generating 20% of its energy from renewable sources by 2020, up from around 8.5%. And the United States aims to produce 20% of its electricity from wind by 2030, up from about 1% today. During the next 10–15 years, Britain is expected to generate more than 30% of its electricity from wind and marine energy sources such as wave and tidal power and to employ between 23,000 and 57,000 people in the wind sector by 2020, up from some 5,000 wind-related jobs in 2008.

Second wind

Europe has long been the global wind-power leader, but much of the recent expansion has been in the United States and Asia. Five years ago, Europe accounted for some 70% of new installations. Now, Europe, North America and Asia rank neck and neck, thanks to strong government policies and large renewable-energy investments. Although countries such as Denmark and Germany still supply the lion's share of wind-turbine parts to Asia and North America, companies are increasingly favouring domestic manufacture of larger and expensive-to-ship parts.

The result is an infusion of new jobs into the United States and Asia. Denmark-based Vestas, the world's largest wind-power manufacturer based on installed megawatts and turbines, opened its first US turbine-blade factory in Windsor, Colorado, in March 2008. It has just broken ground on two more manufacturing plants in Brighton, Colorado.

Wally Lafferty, Vestas vice-president for technology R&D in the Americas, says that the plants should be fully

operational in 2010, and will have manufacturing-related engineering jobs, mainly suited to those educated to bachelor's level with a few posts at master's and PhD level. A Vestas research centre in Houston, Texas, set to open this year, will have 100 positions at master's and PhD levels, for aeromechanical and aerodynamics specialists, high-voltage power transmission and storage specialists, and wind-turbine



"Demand is going to continue to grow."
— Christian Kjaer



Icy blast: wind power in Evanston, Wyoming.

control designers. The company expects to employ some 4,000 people across the United States, including 2,500 in Colorado, by the end of 2010. Late in 2008, Vestas also opened a research centre in Singapore, which now provides more than 100 research jobs for engineers at the master's and PhD levels.

In the United States, where many of the most suitable sites on land are already in use, engineers are working to design turbines for sites with lower wind speeds that will still generate high electrical output, says Greg Watson, vice-president for sustainable development with the Massachusetts Technology Collaborative in Westborough. This includes designing generators with fewer moveable parts and longer, more robust blades that can capture more wind — challenges that will require expertise at the master's and PhD levels, says Watson.

The siting of wind farms now involves more sophisticated mapping technologies and understanding of migratory bird paths than in the past, involving collaborations with biologists and other specialists to assess environmental impacts. And as the industry expands into unexplored regions, experts say there will be a greater need for specialized knowledge from meteorologists and software engineers. They will be required to assess how wind moves over complex terrain such as mountains and seas, to ensure that turbines extract as much wind as possible.

Offshore opportunities

A big portion of future R&D opportunities will be in offshore wind development. The US Department of Energy projects that by 2025, some 70,000 megawatts of capacity could be contributing to the energy grid. Shorter transmission distances to city centres make offshore sites attractive, says Watson. "But the technological and engineering challenges are really tremendous," he adds.

In the United States, nearly 90% of potential offshore wind sites exist in deep water (30 metres or deeper), which will require the development of floating turbine foundations and new materials to withstand pounding waves. Wind-industry experts predict that much research will go into composite materials, adhesives, nanomaterials, 'smart' materials (that can withstand changing conditions such as temperature, moisture and pH) and metallurgy to perfect the 800 or so components that make up a wind turbine.

"Thinking about how to put those together to reduce costs and increase reliability, particularly in a hostile environment like offshore, is a real engineering challenge," says Watson. Offshore power in Europe hasn't yet had to develop deep-water technologies.

Ultimately, how much offshore wind power is developed in the United States will depend largely on state and federal policy, says Willett Kempton, a policy scientist at the University of Delaware in Newark. Currently,



Taking a break, above, and examining a blade.

federal tax credits for renewable energy are helping to make offshore wind development cost-competitive with power from fossil fuels. Delaware has a contract for a 450-megawatt offshore wind farm; it should begin producing electricity within a couple of years. Kempton estimates that building the facility will yield 460 construction and electrical trade positions, 35 engineering positions and 5 openings for scientists. Operating it will require 5 engineers and 15 scientists, along with mechanics and support staff. By today's standards, it would be the world's largest offshore wind farm in terms of megawatt electricity production. Other states, including Massachusetts, Rhode Island, New Jersey and Virginia, are at various stages of the route towards permission for offshore wind development. Kempton expects that, by 2018, the United States will be installing offshore wind capacity at a rate of 500 megawatts per year and will employ hundreds of scientists in design development, impact assessments, manufacturing and installation.

Training for these jobs is in demand, and new programmes are springing up. Delft University offers master's programmes in areas such as materials engineering or control engineering, with a specialization in wind energy. They appeal to students from abroad with bachelor's degrees in mechanical engineering, for instance, who want to apply their skills to wind power, says Gijs van Kuik, scientific director of the Wind Energy Research Institute at the university. The University of Stuttgart, Germany,

offers similar programmes. The Technical University of Denmark in Copenhagen offers a specialized master's diploma in wind energy.

But to meet the surge in demand for workers, Delft and many other universities are also offering programmes for those in industry who want to learn more about specific applications, such as offshore wind or generator design. Because of the massive need for people, says van Kuik, these 'post-academic' courses are being offered two or three times a year.

In addition, the European Academy for Wind Energy will next year start a series of summer courses targeted at master's- and PhD-level education, to allow for further specialization in areas such as meteorology or aerodynamics. Meanwhile, many technical high schools in Europe are setting up programmes to provide technical training in wind energy below the university level.

Graduates of such courses should have reasonable job prospects despite the downturn. Kjaer notes that the

industry might experience some short-term slowdowns, owing to the increased difficulty of obtaining financing or a loss of enthusiasm among investors. "But we're not seeing any decrease in demand for the technology," Kjaer says. "We are in a carbon- and fuel-constrained world and demand is going to continue to grow in the long term."

Amanda Leigh Mascarelli is a freelance writer in Denver, Colorado.



"Wind could give cheaper energy than fossil fuels."
— Willett Kempton

B. FAULKNER/EWEA

BRITISH WIND ENERGY ASSOCIATION

UNIV. DELAWARE

EvoSoap

Truly personal hygiene.

Elizabeth Farnsworth, Aaron M. Ellison and Nicholas J. Gotelli

So convenient, these little refills! And such pretty packaging — with my name printed right there on the front, no less! And the foil keeps the contents sterile, thankfully. And it's so easy to use. According to the instructions, I just open the ziploc and pour the liquid into my dispensers ... in the shower, by the sink, anywhere I need to wash.

And everyone — the chirpy news people, my overanxious daughter — tells me to wash often, of course, to ward off that latest flu or whatever it is that seems to be bringing so many people down. Could it be that Mr Fulton downstairs met his end because he hardly ever washed? (As far as I could tell — his kitchen was a travesty.) A pity, really. And quite awful, the way he went, not to mention how the rest of us in the building had to sit in quarantine for heaven knows how long (really, how many games of solitaire can one play?). But then, he didn't have an educated daughter who could keep him up to date on all the latest health advances. I am so lucky that Vera takes such good care of me.

She told me about EvoSoap and swears by it. Of course, she felt quite pressured to get started with her own series, once the elementary school started mandating it for the children. Kids can be so contagious, after all. I remember not a day seemed to go by without Vera getting an earache or cold when she was a youngster, poor girl, even though the schools vaccinated practically everybody back then! So I suppose it's safest if everybody protects themselves. We all have to do our part.

It's an interesting idea, really: a disinfectant soap tailored to each person's unique genetic profile. I'm no scientist — Vera's father was, but he went years ago in the last outbreak. But Vera (she takes after him) tells me that the Company uses a sample from each person and extracts the genes — 'DNA', is it? Then somehow they can detect if our genes are making antibodies to the latest bacteria or viruses. So then they

make the soap with the right antibodies in it to make up for our deficiencies.

Nobody's perfect!

It seems those bugs fight back, though, changing their own genes to get around our defences. That's why they've got to make new soaps all the time, to adjust — adapt — to the new types of bugs. My daughter says it's a lot like those daily software updates they send to your computer, always staying one step ahead of those mischievous young people — 'hackies', or some such — who

make new viruses for your computer for their fun. Of course, I don't use a computer — too much

my cheek (lots of DNA there, I guess), to PersonalDNA Inc. I was afraid they'd need blood, but thankfully not. I was reluctant at first — I don't want to send parts of me to some stranger! But those slick TV ads — honestly, they're on every ten minutes! — gave me the impression that these people know what they're doing. They're real professional scientists — smart, just like my Vera. It's a little disturbing to me that they keep my genes on file — or is it just the sequence of letters in my genes ... oh, yes, the 'code' — in that big national data bank. I don't know where they possibly have room for all those codes! But the Homeland Security office knows best; they say they can use the information to catch crooks, perhaps even those hackies. And, at my age, I confess I don't really care; I'll be leaving this Earth soon anyway. And actually, I haven't had so much as a sniffle in years.

It's a little cumbersome, as they send me a new 'update' soap every month. I guess it's no more trouble than a magazine subscription, but the monthly payments and the express mail costs do take a bite out of my budget. I've been on fixed income since Jeremy died, but Vera helps out. And, as I said, we all have to do our part.

The president told us as much on his daily TV address about the outbreak. Silly man — he calls it the 'War on Germs'. I remember a decade ago — or was it two? — when that other president had his very own 'War on Terror' and somebody before that had a 'War on Poverty'. As if every mother in America, including me and Vera, weren't waging our own little wars on germs with our children. But this president is too young to remember polio. And now he's so hot-up about this flu, telling us to wash, wash, wash using EvoSoap. If I didn't know better, I'd say that the Company is running this country!

Oh, there's the doorbell. Must be the mailman with a new packet for me — I was almost out of the old one! He's such a sweet boy — I've got to know him so well. But I worry about that little cough of his. ■

Elizabeth Farnsworth, Aaron Ellison and Nicholas Gotelli — scientists all — came up with this idea over breakfast (which we ate only after washing our hands).



trouble to learn at my age! I never was a big believer in that so-called 'evolution' theory, either, but my daughter tells me 'mutation' is a big problem, so I've got to use the new soap to keep up with the changes, as she says the bugs can develop resistance pretty fast. I wish I could still use my old soap, but I guess it doesn't work anymore, and besides, there's only one soap on the market.

So, on her insistence — she's so lovely, Vera, but always so worried — I sent in my sample, a Q-tip swab of the inside of

JACEY